PROCEEDINGS
11th Islamic Countries Conference on Statistical Sciences (ICCS-11)
Statistics for Strategic Development

on
December 19-22, 2011

at
University of Management and Technology
Lahore, Pakistan

Jointly organized by

Islamic Countries Society of Statistical Sciences
National College of Business Administration & Economics
University of Management and Technology

Islamic Countries Society of Statistical Sciences
44-A, Civic Centre, Sabzazar, Multan Road, Lahore, Pakistan (www.isoss.net)
“All papers published in the PROCEEDINGS of ICCS-11 were accepted after formal peer review by the experts in the relevant field.

Dr. Munir Ahmad
Editor
<table>
<thead>
<tr>
<th>Contents</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Address by Dr. Munir Ahmad, Founding President and Patron ISOSS</td>
<td>vii</td>
</tr>
<tr>
<td>1. 006: ARMA mortality regression module of lung cancer in Palestine</td>
<td>1-10</td>
</tr>
<tr>
<td>Hazem I. A. El Shekh Ahmed</td>
<td></td>
</tr>
<tr>
<td>2. 009: Computer science algorithms in space science</td>
<td>11-14</td>
</tr>
<tr>
<td>Irshad Ullah</td>
<td></td>
</tr>
<tr>
<td>3. 010: Data quality enhancement using data mining approach</td>
<td>15-18</td>
</tr>
<tr>
<td>Irshad Ullah</td>
<td></td>
</tr>
<tr>
<td>4. 028: Testing of intercept when slope is under suspicion</td>
<td>19-29</td>
</tr>
<tr>
<td>Budi Pratikno and Shahjahan Khan</td>
<td></td>
</tr>
<tr>
<td>5. 034: Trends in research publications of Pakistani statistical journals</td>
<td>31-38</td>
</tr>
<tr>
<td>Mehwish Hussain and Waqas Ahmed Farooqui</td>
<td></td>
</tr>
<tr>
<td>6. 038: Instrumental variable estimator of the slope parameter when the explanatory variable is subject to measurement error</td>
<td>39-53</td>
</tr>
<tr>
<td>Anwar Saqr and Shahjahan Khan</td>
<td></td>
</tr>
<tr>
<td>7. 041: Bayesian prediction of matrix elliptical multivariate models with conjugate prior</td>
<td>55-65</td>
</tr>
<tr>
<td>Mohammad Arashi and Shahjahan Khan</td>
<td></td>
</tr>
<tr>
<td>8. 047: Modelling the daily rainfall amounts of Dir and D.I. Khan, Pakistan for agricultural planning</td>
<td>67-82</td>
</tr>
<tr>
<td>Zahid Hussain</td>
<td></td>
</tr>
<tr>
<td>9. 051: Non-parametric and parametric duration analysis of coronary artery disease with chest pain in past</td>
<td>83-90</td>
</tr>
<tr>
<td>Mehwish Hussain, Nazeer Khan and Mudassir Uddin</td>
<td></td>
</tr>
<tr>
<td>10. 053: A brief review of geoinformatics analysis on poverty data in Indonesia</td>
<td>91-102</td>
</tr>
<tr>
<td>Asep Saefuddin</td>
<td></td>
</tr>
<tr>
<td>11. 055: Relationship among GDP, per capita GDP, Literacy rate and unemployment rate</td>
<td>103-111</td>
</tr>
<tr>
<td>M. Shafiqur Rahman</td>
<td></td>
</tr>
<tr>
<td>12. 058: The consumption pattern of household food-away-from-home: Pakistan</td>
<td>113-118</td>
</tr>
<tr>
<td>Ammara Nawaz Cheema and Naheeda Akhtar</td>
<td></td>
</tr>
<tr>
<td>13. 059: Redundant use of mobile affects student’s GPA</td>
<td>119-126</td>
</tr>
<tr>
<td>Ammara Nawaz Cheema</td>
<td></td>
</tr>
<tr>
<td>Waleed Ameen Abd Elkhalek Mohamed</td>
<td></td>
</tr>
<tr>
<td>15. 119: A proposed approach on multivariate clustering analysis using simulated annealing - A case study on location problem</td>
<td>143-150</td>
</tr>
<tr>
<td>Mahdi Bashiri and Mohammad Khanloo</td>
<td></td>
</tr>
<tr>
<td>16. 120: Statistical analysis of the number of the galaxies in cubic cells in the universe</td>
<td>151-160</td>
</tr>
<tr>
<td>Fahimeh Mostajeran</td>
<td></td>
</tr>
<tr>
<td>17. 125: Universally optimal neighbor designs in blocks of size three</td>
<td>161-172</td>
</tr>
<tr>
<td>Munir Akhtar and Rashid Ahmed</td>
<td></td>
</tr>
<tr>
<td>No.</td>
<td>Title</td>
</tr>
<tr>
<td>-----</td>
<td>----------------------------------------------------------------------</td>
</tr>
<tr>
<td>18.</td>
<td>Recursive method for AR parameters in the presence of outlier in poisson noise</td>
</tr>
<tr>
<td>19.</td>
<td>Bayesian analysis of the adaptive type-II progressively hybrid censoring scheme in presence of competing risks</td>
</tr>
<tr>
<td>20.</td>
<td>Root transformation method for monitoring correlated variable and attribute quality characteristics</td>
</tr>
<tr>
<td>21.</td>
<td>Using time dependent coefficient rates model for assessing the impact of breastfeeding on weight loss in less than two years old children</td>
</tr>
<tr>
<td>23.</td>
<td>A note on bathtub hazard rate</td>
</tr>
<tr>
<td>24.</td>
<td>Mixture ratio and regression estimators using multi-auxiliary variables and attributes in single phase sampling</td>
</tr>
<tr>
<td>25.</td>
<td>Comparison of three different methods for outlier detection in bioequivalence studies</td>
</tr>
<tr>
<td>27.</td>
<td>Quantifying the effects of turbulent atmosphere on small satellites at Pakistan ionospheric environment</td>
</tr>
<tr>
<td>28.</td>
<td>On the role of the multivariate fatigue life distribution in modeling multi-site fatigue</td>
</tr>
<tr>
<td>29.</td>
<td>Economic design of time truncated sampling plans for weibull distribution using percentiles as quality parameter</td>
</tr>
<tr>
<td>30.</td>
<td>Role of line manager in implementing Human resource management and human resource development</td>
</tr>
<tr>
<td>31.</td>
<td>Estimation of the scale and location parameters based upon the upper records values of weibull distribution from complete data, type-II singly and doubly censored data</td>
</tr>
<tr>
<td>32.</td>
<td>Estimation of the parameters based on lower record values from the gumbel distribution for type-II singly and doubly censored data</td>
</tr>
<tr>
<td>33.</td>
<td>Prediction of household expenditure on the basis of household characteristics</td>
</tr>
<tr>
<td>No.</td>
<td>Paper Title</td>
</tr>
<tr>
<td>-----</td>
<td>-------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>34.</td>
<td>M-070: Classification of households with respect to poverty by using cluster analysis</td>
</tr>
<tr>
<td>36.</td>
<td>M-074: Classification of countries with respect to their governance</td>
</tr>
<tr>
<td>37.</td>
<td>M-075: Problems in the use of educational technology</td>
</tr>
<tr>
<td>38.</td>
<td>M-076: Computer hacking forensics investigation techniques</td>
</tr>
<tr>
<td>40.</td>
<td>M-080: Techniques used for hacking the information</td>
</tr>
<tr>
<td>41.</td>
<td>M-081: On selection of autoregressive order in case of incorrectly model specification</td>
</tr>
<tr>
<td>42.</td>
<td>M-082: New variational model for images elective segmentation</td>
</tr>
<tr>
<td>43.</td>
<td>M-087: A concept for personage’s learning style centered technology based adaptive pedagogical system</td>
</tr>
<tr>
<td>44.</td>
<td>M-088: Cognitive model for social bots to learn opponents mixed strategy using adaptive neural networks</td>
</tr>
<tr>
<td>45.</td>
<td>M-089: SKANS: secure knots at the ends</td>
</tr>
<tr>
<td>46.</td>
<td>M-090: Building conscious cybernetic entities using QuBIC model and framework based on unified theory of mind</td>
</tr>
<tr>
<td>47.</td>
<td>M-091: Determinants of food inflation in pakistan &amp; effects of seasonal adjustment on forecasting food inflation</td>
</tr>
<tr>
<td>49.</td>
<td>M-097: Underlying issues in knowledge elicitation towards automation</td>
</tr>
<tr>
<td>51.</td>
<td>M-100: An improved class of regression estimators for the population mean using multi-phase sampling scheme in the presence of non-response</td>
</tr>
<tr>
<td>No.</td>
<td>Title</td>
</tr>
<tr>
<td>-----</td>
<td>----------------------------------------------------------------------</td>
</tr>
<tr>
<td>52</td>
<td>M-101: A general family of estimators when the study variable is an attribute</td>
</tr>
<tr>
<td>53</td>
<td>M-102: Estimation of finite population coefficient of variation in stratified random sampling</td>
</tr>
<tr>
<td>55</td>
<td>M-107: Statistical study of agriculture development in Pakistan</td>
</tr>
<tr>
<td>56</td>
<td>M-114: A quantum inspired sparse distributed memory (QI-SDM) model for clean patterns retrieval</td>
</tr>
<tr>
<td>58</td>
<td>M-116: Role of emotion oriented agents in information media for effective communication</td>
</tr>
<tr>
<td>59</td>
<td>M-119: Performance excellence through integration of KM philosophy in TQM practices</td>
</tr>
<tr>
<td>60</td>
<td>M-121: GIS framework using SDI for disaster management</td>
</tr>
<tr>
<td>62</td>
<td>M-123: Circulating serum concentrations of adipocytokines in obese and type 1 diabetic children and young adolescents</td>
</tr>
<tr>
<td>63</td>
<td>M-124: Operating systems scheduling a new perspective</td>
</tr>
<tr>
<td>64</td>
<td>M-126: Sudoku square design, layout and analysis</td>
</tr>
<tr>
<td>65</td>
<td>M-127: Statistical meta-analysis for ordinal categorical data</td>
</tr>
<tr>
<td>66</td>
<td>M-129: Factor analysis of teacher’s intervention strategies for solving student’s behavior and learning problems</td>
</tr>
<tr>
<td>67</td>
<td>M-135: Wireless security and threats</td>
</tr>
<tr>
<td>68</td>
<td>M-143: Reifying techno-e learning as catalyst for affirmative change</td>
</tr>
<tr>
<td>69</td>
<td>N-1: Weighted estimators of population mean using two auxiliary variables under two phase sampling</td>
</tr>
</tbody>
</table>
Address by Dr. Munir Ahmad  
Founding President and Patron ISOSS

- Prof. Dr. Ali S. Hadi, President-Elect ISOSS (2011-2013)  
- Dr. Shahjahan Khan, President ISOSS  
- Professor Dr. Hassan Sohaib Murad,  
  Rector, University of Management & Technology, Lahore  
- Dr. Abdul Raouf, Prof. Abdus Samad Hirai, Dr. A.Z. Memon, Dr. M. Hanif Mian  
- My colleagues and my students  
- Ladies and Gentlemen!

I on my behalf and on behalf of members of ISOSS welcome you to the 11th Islamic Countries Conference on Statistical Sciences especially those who come from Australia, Indonesia, Malaysia, Sudan, Oman, Egypt, Iran, Canada, USA and UK.

First of all I wish to thank Dr. Shahjahan Khan, President ISOSS for his exemplary leadership of ISOSS for the last 6 years. The Society, under his visionary leadership had been very successful and he worked untiringly and steward the Society’s ship to its present position. I stand witnessed to the impressive progress ISOSS has made during his tenure. The momentum of our efforts with his enthusiasm has tremendously increased.

Second, I would like to thank Prof. Dr. Ali S. Hadi for acceptance of the result of his election as President-Elect of the Society. We look forward to his dynamic stewardship of the Society, which he demonstrated by organizing 10th Islamic Countries Conference on Statistical Sciences at Cairo, Egypt. I am sure, ISOSS will take new shapes and directions during his tenure-ship. I assure you of our full support.

I also like to thank Professor Dr. Hassan Sohaib Murad, Rector, University of Management and Technology (UMT), Lahore for sponsoring the conference, who for the love of education has dedicated himself to the spreading of knowledge. I deeply appreciate his efforts for UMT programs by creating alliances amongst scientists of various disciplines and encouraging us in organizing this conference for exchange of knowledge and experiences for better understanding and advancement of science.

I feel proud that our efforts have become fruitful in holding the 11th Islamic Countries Conference on Statistical Sciences when there are many socio-political changes being held in the Islamic World.

The society, which had celebrated its 23 years journey in 2011, was established way back in 1988, during the first Islamic Countries Conference on Statistical Sciences held at Lahore.

I feel really proud of its existence, as the Society had held ten Islamic Countries Conferences on Statistical Sciences and numerous National Conferences, Seminars and Workshops in Islamic Countries.
The core tool of research is statistics and computer is the tool for statistics. Statistics-based computer technology is the main stay of scientific advancement. However, statistics has to become self-reliant. Most of jobs in all sectors have more statistics content than contents of the jobs itself. All workers or what I call “PARASTATISTICIANS” have to have a fair knowledge of basic statistics and have to have proper statistics language and to understand the usefulness of computer software packages.

I strongly believe that planning has to be knowledge based and to be monitored by a strong team of statisticians. At present, there is no interaction between academia and Statistics Officials. If Official Statistics is to be improved, academia must be involved very effectively and if statistical agencies need qualified incumbents, statistical agencies and industries must be effectively involved in colleges and universities so that statisticians produced by colleges and universities are directly inducted in statistical organizations and industries.

I foresee ISOSS would develop into a World Forum that can be managed on a collective vision of its active members. Dignity of top statisticians working both in public and private sectors is a pre-requisite to Society’s strategy and action plans.

You will be glad to know that ISOSS has now built its own ISOSS House. Members and friends of ISOSS had generously contributed to its construction work and are contributing / finances for the ISOSS House. We still need help from all friends of Statistics for setting up internationally recognized Research Institution and equipping with necessary facilities. ISOSS activities have been internationally recognized. In this connection I also thank (Mian Shahid Ali Haider, Vice-Chairman), NCBA&E for full support to make ISOSS activities fully operational.

We need to look up to our ancestral scientific legacy. We cannot revive our scientific heritage by putting education outside the framework of national activities and seeking assistance from outside Islam. We need program with faith and follow the foot steps of our administrative giants. Hazrat Omar (RS), bravery and intellect of Hazrat Ali (AS). We had created and established many great educational institutions in the past but we could not sustain those great universities. We let others destroy our all scientific and academic institutions and libraries which were rich and well established in Baghdad, Cordoba, Seville, Cairo, Toledo, and Fez. The calamity of oblivion of Islamic thoughts still persists. In the past our Ulemas were scientists and our scientists were renowned Ulemas. They were Imams, scientists and poets, who produced un-precedent inventions, discoveries, and philosophy. Those are still the bases of current scientific advancements in the world. We were no doubt proud of Muslim scientists and their philosophic thoughts. Even now we have noble laureates in the Muslim world which show that we have the potential of becoming great scientists, social scientists and philosophers. We should not forget our past and current achievements but unfortunately, we still seem to live in Mohanjo Daro and pre-Egyptian civilization. It is still a common sight of wooden bullock driven carts on broken roads alongside Pajero and BMW running on Motor Ways in most of the Islamic Countries. We are still proud of our ancestors and take pride in what they had achieved in the fields of science and technology. They excelled others in their achievements in many aspects of contemporary sciences. They kindled the lights of
knowledge around the globe and shared unreservedly the fruit of their learning with others. It paints of very dismal and gloomy picture when we look at our present status and the directions of which we are heading today. I am not talking about the Muslim scientists; I am talking about the overall status of science and technology in the Muslim World. In the past, other nations learnt many things from us. Now, we will have to learn from them in order to make advancements in the current state of our knowledge at an accelerated pace as a religious obligation. With committed and devoted applications, we can recreate, under the new system of coordination and interaction among Muslim scientists and scientific centers of Fez, Azhar, Maragha, Iran, Cordova, Grenada, Baghdad, Basra, Tashkent, Bukhara and Samarkand. These were the places where scientists from both East and West were attracted to. At this point, in our history, the Muslim communities have produced Scientists who were capable of excelling in their areas of specializations and were able to contribute significantly. However, this is very unfortunate that Muslim Ummah is unable to utilize the potential of their capable scholars and scientists. Unfortunately these persons have been serving the Western societies. It is an economic reality that wealth migrates to safe abodes. The same is true for the scientists and technologists who are attracted towards seats of learning and places of appreciation. They migrate to places where they are secure, welcomed, honoured and nurtured. I am sure you will agree with me that this is an opportune time to accept the challenges of rediscovering the glory of past and leading scientists to enter the 21st Century with prestige, honour and respect. It is a common question, “how long will it take us to revive our scientific and cultural supremacy?” I believe, if we take faith like our ancestors and act with their dedication and commitment, certainly, it will not take us too long to achieve the same heights of the past glory.

In the end, I thank my team of volunteers, mostly students and professors (from National College of Business Administration and Economics), especially Dr. Muhammad Hanif, Mr. Anwer Hasnain, Dr. M. Moeen, Haider and the names I do not remember and the most important of all is my ISOSS Secretariat staff Muhammad Iftikhar, Muhammad Imtiaz and Saif-ur-Rehman and to others for their untiring work. There is a long list of students and persons who made this conference a success. I am especially proud of work done by Dr. Faisal Qadeer, and Dr. Alia Ahmed and their team members and the members of the Conference Committee at UMT especially Mr. Noor Aslam, Operational Manager and his team at UMT. He and his team has worked very hard for the conference arrangements. I am extremely grateful to Dr. Hasan Sohaib Murad, Rector, UMT for the facilities provided to the conference.

I again show my gratitude to Dr. Shahjahan Khan and Dr. Ali S. Hadi and delegates from Australia, Indonesia, Malaysia, Sudan, Egypt, Iran, Canada, USA, UK and Oman. I thank you Sir, and thank you all.
ARIMA MORTALITY REGRESSION MODULE OF LUNG CANCER IN PALESTINE

Hazem I. El Shekh Ahmed
Al-Quds Open University, Khanyounis Educational Region, Gaza Strip, Palestine.
Email: hshaikahmad@qou.edu

ABSTRACT

BACKGROUND: Worldwide, Lung Cancer is the most commonly diagnosed cancer and causes more deaths than any other one. Its high mortality rate results from both a high incidence and a low survival rate. Lung Cancer diseases are considered one of the most serious health problems in the developed and developing countries for their high cost, incidences and mortality. We have used cancer registry data in the Palestinian Ministry of Health associated with mortality rates in the West Bank and Gaza Strip.

METHOD: We apply the techniques of the time series analysis from Box, et al. (1994) in two time series, Series A and Series B depending on the reported motility data in the Health Centers Registry (Reports of the Palestinian Ministry of Health, 1995-2005).

RESULTS: The mortality rate of lung cancer, \( z_t \) denotes the value of the time series at time \( t \), and \( \nabla z_t = z_t - z_{t-1} \) denotes its first difference at time \( t \). The researcher began the previous time-series analysis using S-Plus program through fitting a smooth chart to identify the general pattern of mortality rates during the time period covered by the study in the Palestinian territories. Because of a white noise might appeared, we tried some different models using the S-Plus program and the best one was the one that described by an ARIMA(1,1,0) model since it has the smallest Akaike information criterion (AIC=0.02093). The fitted model for Series A is given by: \( \nabla z_t - 0.41 \nabla z_{t-1} = \epsilon_t \) with standard error (± 0.41) of the estimated coefficient, and \( \epsilon_t \) is a white noise. We predict the rates of mortality in Palestinian Territories for the next 5 years (2006 – 2011), and they were 4.46, 4.53, 4.56, 4.57, and 4.58 per 100 000 respectively.

CONCLUSION: All available evidence from this study indicates that the increasing of expected mortality rates resulted by the ARIMA module must be attended and faced. The risk factors of lung cancer can be studied locally in Palestine and then make the respected limitation in these increasing rates.

KEYWORDS

Lung cancer; ARIMA regression; fitted module; mortality rates.

INTRODUCTION:

The study of population health is focused on understanding health and disease in community, and on improving health and well-being through priority health approaches addressing the disparities in health status for some dangerous diseases in order to have a
clear plan to face such cases. Cancer diseases are considered one of the most important health problems in the developed and developing countries for its high cost, incidences and mortality. Lung cancer is a major global health burden with high incidence rates but poor long-term survival. Lung cancer is a major worldwide health burden, responsible for 1.3 million deaths in 2004, equating to 2.3% of all deaths. Death rates from lung cancer are predicted to continue to rise, with the disease being responsible for 2.8% of all deaths (1.67 million) by 2015 (ATHEY et al. 2010). Several studies have interested with the risk factor of the lung cancer Despite advances in treatment and diagnosis, the statistical tools are so vital to Struggle this killer disease. The ARIMA regression module can produce a suggested form related with time and generate expected values.

Locally, the total number of Palestinian people according to the estimation 2009 was 3,935,249 of which 50.8% are males and 49.2% are females. Cancer diseases were the third leading cause of death in the Palestinian territories. There was a slight increase in Cancer mortality in 2009 compared with 2007, from 10.5% to 11% from the total deaths. Trachea cancer was the highest one; it was 17.3 %, Colo-rectal and anus Cancer 13.3 %. Cancer mortality in Males is higher than in Females (56.9 %) & (43.1 %). (Health Annual Report, 2009).

This paper provides an overview of the current evidence base regarding mortality cases of lung cancer in the Palestinian Authority (Gaza Strip and West Bank) and suggests some regression module describing the expected mortality cases in the future.

**METHODOLOGY**

This part of the article provides a brief introduction to the major concepts and terminology of ARIMA methodology, discusses the need for model identification within the scope of process analysis, introduces automated methods for model identification, and reviews past studies evaluating the performance of automated procedures.

A time series is a sequence of values ordered by a time parameter. ARIMA represents a general class of time series models that combines several time series techniques such as differencing, autoregressive models, and moving average models. Each time series can be described by means of three types of mathematical models: autoregressive (AR), moving-average (MA), and integrated (I).

An autoregressive (AR) model of the order p is the one in which the current observation \( z_t \) is occurred in some previous observations \( z_{t-1}, z_{t-2}, ..., z_{t-p} \) of the same time series, we can expressed this as:

\[
   z_t = \xi + \varphi_1 z_{t-1} + \varphi_2 z_{t-2} + ... + \varphi_p z_{t-p} + e_t
\]

where \( \varphi_1, \varphi_2, ..., \varphi_p \) are the regression coefficients, \( \xi \) is the constant term, and \( e_t \) is the random error term. The term \( e_t \) is an independently, identically distributed (i.i.d.), zero-mean sequence of random shocks with variance \( \sigma_e^2 \).

By direct calculations eq.(1) can be rewritten as,

\[
   z_t - \varphi_1 z_{t-1} - \varphi_2 z_{t-2} - ... - \varphi_p z_{t-p} = \xi + e_t
\]
Let $B$ denote the backward shift operator such that, $Bz_t = z_{t-1}$, $Bz_{t-1} = z_{t-2}$, and so on, then eq.(2) can be expressed as,

$$
(1-\varphi_1 B - \varphi_2 B^2 - \ldots - \varphi_p B^p)z_t = \xi + e_t
$$

(3)

now, if $z_t = z_t - \mu$, then eq.(3) becomes

$$
\varphi_p (B)z_t = e_t
$$

(4)

where, $\varphi_p (B) = (1 - \varphi_1 B - \varphi_2 B^2 - \ldots - \varphi_p B^p)$

On the other hand the moving average (MA) model of order $q$ can be expressed in terms of the random errors terms as,

$$
z_t = \mu + e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \ldots - \theta_q e_{t-q}
$$

(5)

Again using the backward shift operator $B$, eq.(5) can be represented as,

$$
z_t = \mu + \left(1 - \theta_1 B - \theta_2 B^2 - \ldots - \theta_q B^q\right)e_t
$$

(6)

or briefly,

$$
z_t = \mu + \theta_q (B)e_t
$$

(7)

where, $\theta_q (B) = 1 - \theta_1 B - \theta_2 B^2 - \ldots - \theta_p B^p$.

The two models, autoregressive AR($p$) and moving average MA($q$) can be combined to form a mixed Autoregressive-Moving Average model (ARMA($p,q$)) expressed as the equation,

$$
\varphi_p (B)z_t = \xi + \theta_q (B)e_t
$$

(8)

An equivalent form of eq.(8) can be derived and written as, (Box et al. 2007)

$$
\varphi (B)z_t = \theta (B)e_t
$$

(9)

where $e_t$ is the white noise.

The previous time series is supposed to be stationary: the statistical distribution of a set of random variables is said to be strictly stationary if the joint probability distribution at any set of times $t_1, t_2, \ldots, t_m$ is the same as the joint probability distribution at times $t_{1+k}, t_{2+k}, \ldots, t_{m+k}$ for all integers $m$ and $k$ (Krishan Kumar, 2004). If the time series is nonstationary, it can be transformed into a stationary time series using a process called differencing in which the previous observation $z_{t-1}$ is subtracted from the current observation $z_t$. The difference operator $\nabla$ will be used to represent the differencing between time series terms, and will be defined as, $\nabla z_t = z_{t-1} - z_t = (1-B)z_t$. 

differencing of order 1, where B is the backward shift operator. It is clear that
\[ B(z_t) = z_{t-1} \text{ and } \nabla = (1 - B), \] generally \( \nabla^k z_t = \nabla \left( \nabla^{k-1} z_t \right) \) is a differencing of order k.

The first-order difference operator can eliminate linear trend component, and consequently the \( n^{th} \)-order difference operator aims at eliminating \( n^{th} \)-order trend component (M. Zhou, et al. 2006). If there is no periodic component in \( P(t) \), then the difference operator \( \nabla \) can be applied repeatedly until a stationary process is arrived. The first and second order is commonly used in the practical cases to obtain the stationary time series.

Combining the three techniques (autoregressive models, moving average models, and differencing), an ARIMA model is generated. ARIMA model is represented by ARIMA \((p, d, q)\), where \( p \) is the order of the autoregressive part, \( d \) is the order of differencing and \( q \) is the order of moving-average process. A time series \( z_1, z_2, ..., z_k \) generated by an ARIMA \((p, d, q)\) process with mean \( l \) of the Box–Jenkins model (Box G.E.P. and Jenkins, G.M., 1976) is expressed as follows:

\[
\phi(B)(1-B)^d (z_t - \mu) = \theta(B)e_t
\]

Eq.(10) can be rewritten as,

\[
\phi(B)\nabla^d (z_t) = \theta(B)e_t
\]

In the present study, \( z_t \) and \( e_t \) represent mortality rate of lung cancer incidences in Palestine and random error terms at time \( t \), respectively, while \( d \) represents the order of differencing.

The Box – Jenik ARIMA modeling approach procedure consists of a four step iterative procedure.

1) model identification, where the orders \( d, p, q \), are determined by observing the behavior of the corresponding autocorrelations and partial autocorrelations; 2) estimation, where the parameters \( \phi_1, \phi_2, ..., \phi_p, \theta_1, \theta_2, ..., \theta_q \) of the model are to be estimated; 3) diagnostic checking, where the adequacy of the fitted model is checked; 4) making a decision, that’s make by the result of stage 3. If the model is judged be inadequate, the stages 1 – 3 are repeated with different values \( d, p \) and \( q \) until an adequate model is obtained (Ozaki, 1977).

**DATA DESCRIPTION**

The mortality data of lung cancer patients (for the whole population per 100000 residents) provided in this paper are taken from the Palestinian Ministry of Health annual reports from 1995 to 2005. Two mortality data groups in different period was obtained, the first mortality rates of Palestinian Territories, have been considered for the period 1999 –2005, and the second group restricted in Gaza strip during 1995 to 2005.

The first step towards identifying ARIMA tentative model passes through the analysis of historical data in the two groups.
This implies examining whether the time series is stationary or not. The first time series data A of lung cancer mortality rates in Palestinian Territories is plotted using S-Plus program, see Fig.(1)

Fig.1: Rates of lung cancer Mortality in Palestine Territories (1999 – 2005).

The second time series B of lung cancer mortality rates in Gaza Strip is also plotted using S-Plus program, see Fig.(2)

Fig. 2: Rates of lung cancer Mortality in Gaza Strip (1995 – 2005).
The previous two time series A and B were nonstationary white noise so it must be treated by identifying the ARIMA constants p, d and q using S-Plus program.

**RESULTS AND DISCUSSIONS**

The estimated autocorrelations and partial autocorrelations of series A and its first difference are plotted in Figure 3.

**Fig. 3: The estimated autocorrelations and partial autocorrelations of series A and its first difference.**

These plots suggest that this time series might be a white noise, so we tried some different models using the S-Plus program and the best one was the one that described by an ARIMA(1,1,0) model since it has the smallest Akaike information criterion (AIC=0.02093). The fitted model for Series A is given by:

\[
\nabla z_t - 0.41 \nabla z_{t-1} = e_t
\]

with standard error (± 0.41) of the estimated coefficient, and \( e_t \) is a white noise.

Equation 12 represents the ARIMA (1,1,0) model which used to predict the rates of mortality in Palestine Territories for the next 5 years and they were 4.46, 4.53, 4.56, 4.57, and 4.58 respectively starting from 2006.

Plot of series A together with our predictions for the next five years are shown in Figure 4.
Fig. 4: The predicted mortality rates of lung cancer in Palestinian territories (2006-2010).

Later, the available rates of mortality for lung cancer in Gaza Strip during the period 1995 – 2005 were analyzed by using the time series B, and fitting a smooth chart to identify the general pattern of mortality rates during the time period covered by the study in the Gaza Strip (see Fig.2).

Again, the estimated autocorrelations and partial autocorrelations of series B and its first difference are plotted in Figure 5. These plots suggest that this time series might be a white noise, so we tried some different models using the S-Plus program and the best one was the one that described by an ARIMA(2,1,0) model. Using the S-Plus program, the fitted model for series B is given by:

$$\nabla z_t + 0.58 \nabla z_{t-1} + 0.69 \nabla z_{t-2} = e_t$$  \hspace{1cm} (13)

with the same standard errors (± 0.256 ) and (± 0.256 ) of these estimates respectively, and $e_t$ is a white noise.

Fig. 5: The estimated autocorrelations and partial autocorrelations of series B and its first difference.
Equation 13 represents the model used to predict the rates of mortality in Gaza Strip for the next 5 years and they were 3.92, 3.78, 4.19, 4.05 and 3.85 starting with 2006, respectively.

Plot of series B together with our predictions for the next five years are shown in Figure 6.

Fig. 6: The predicted mortality rates of lung cancer in Gaza Strip(2006-2010).

The previous partial results more predictable than previous rates, based on statistical models of former (12) and (13). Knowing that the degree of accuracy of such models is to increase the raw data available.

MODEL VALIDATION

In addition to investigating the goodness of fit of a model, it is equally important to validate the model forecasts. Because of the forecast errors of future times are not known, ARIMA models are widely applied and possess various good qualities. Thus, in the present study, the model validation was done by obtaining one-period-ahead forecasts for the last 11 observations of the time series A and 6 observations of the time series B, which were not a part of the initial model fitting. After every forecast, the forecast error (forecasted value - observed value) was computed and the observed value became part of the data for model fitting for the next prediction in the two cases.

The mean absolute percentage error (MAPE) for the forecasts of series A is found to be about 22.15%, and 2.19% for series B (Table 1), which is reasonably good, given the stochastic nature of the mode. This value is expected to be reduced further when the model is trained with a longer time series of observed data. Some of the most commonly used performance measures for the evaluation of mortality rates models are Root Mean Square Error (RMSE), and Normalized Bayesian Information Criterion (NBIC). The computed values of these measures are presented in Table 1. The values reveal that the model’s performance is quite satisfactory.
Table 1

<table>
<thead>
<tr>
<th>Evaluation parameters of model A performance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model parameters</td>
</tr>
<tr>
<td>$\phi_1$</td>
</tr>
<tr>
<td>Standard Error (SE)</td>
</tr>
<tr>
<td>Mean Absolute Error (MAE)</td>
</tr>
<tr>
<td>Akaike Information Criterion (AIC)</td>
</tr>
<tr>
<td>Mean Absolute Percentage Error (MAPE)</td>
</tr>
<tr>
<td>Root Mean Square Error (RMSE)</td>
</tr>
<tr>
<td>Normalized Bayesian Information Criterion (NBIC)</td>
</tr>
</tbody>
</table>

Table 2 describes the parameters of ARIMA model B performance, there was two autoregressive coefficients $\phi_1$, $\phi_2$ with values 0.58, 0.69 respectively. The standard error is found to be about ± 0.256, ± 0.256. More computed measures of the time series B are presented in table 2 and these values indicate model’s performance is quite satisfactory.

Table 2:

<table>
<thead>
<tr>
<th>Evaluation parameters of model B performance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model parameters</td>
</tr>
<tr>
<td>$\phi_1$, $\phi_2$</td>
</tr>
<tr>
<td>Standard Error (SE)</td>
</tr>
<tr>
<td>Mean Absolute Error (MAE)</td>
</tr>
<tr>
<td>Akaike Information Criterion (AIC)</td>
</tr>
<tr>
<td>Mean Absolute Percentage Error (MAPE)</td>
</tr>
<tr>
<td>Root Mean Square Error (RMSE)</td>
</tr>
<tr>
<td>Normalized Bayesian Information Criterion (NBIC)</td>
</tr>
</tbody>
</table>

CONCLUSION

An ARIMA (1,1,0) model of time series A that describes the mortality of lung cancer for the next five years in Palestine will increase slightly comparing with the level in the past years. Values of the model coefficients and patterns of ACF and PACF in fig. 3 indicate the dominance of the Autoregressive Process of order 1 in the model.

We see from these results related with time series B of ARIMA (2,1,0) describes the mortality of lung cancer for the next five years in Gaza will be around the same level in the past years. Also the values of the model coefficients and patterns of ACF and PACF in fig. 5 indicate the dominance of the Autoregressive Process of order 1 in the model. The accuracy of forecasted values to the observed values of the mortality rates was reasonable closed with MAPE 22% and 2% respectively. Computed values in table 1,2 of the two model performance measures further confirm the model’s forecast capability.

This study confirms the effectiveness of ARIMA models for obtaining short term forecasts of mortality lung cancer, and that’s leads to benefit dependence in the previous results in the future health planning strategies.
REFERENCES


COMPUTER SCIENCE ALGORITHMS IN SPACE SCIENCE

Irshad Ullah
Department of Computer Science, Lahore University of Management Science, Lahore, Pakistan
Email: irshadullah79@gmail.com

ABSTRACT

Science gives mankind motivation and hope. Space science makes us look outwards from our planet, towards the stars. By studying alien worlds, such as Venus, Mars, or Saturn’s moon Titan, we can put our own in context. In this paper data mining approach will be applied to the data from space science. The outcome will be helpful for the space scientist. These results may be helpful to minimize discrimination in the said field.

KEY WORDS

Space; Planet; Algorithm; Mankind; Data.

1. INTRODUCTION

1.1 Space Science

Space science is an all-encompassing term that describes all of the various science fields that are concerned with the study of the Universe, generally also meaning "excluding the Earth" and "outside of the Earth's atmosphere". Originally, all of these fields were considered part of astronomy. However, in recent years the major sub-fields within astronomy, such as astrophysics, have grown so large that they are now considered separate fields on their own. There are eight overall categories that can generally be described on their own; Astrophysics, Galactic Science, Stellar Science, non-Earth Planetary Science, Biology of Other Planets, Astronautics/Space Travel, Space Colonization and Space Defense [5].

1.2 Data Mining

(the analysis step of the Knowledge Discovery in Databases process[3], or KDD), a relatively young and interdisciplinary field of computer science, [4][1] is the process of discovering new patterns from large data sets involving methods from statistics and artificial intelligence but also database management. In contrast to for example machine learning, the emphasis lies on the discovery of previously unknown patterns as opposed to generalizing known patterns to new data.

The term is a buzzword, and is frequently misused to mean any form of large scale data or information processing (collection, extraction, warehousing, analysis and statistics) but also generalized to any kind of computer decision support system including artificial intelligence, machine learning and business intelligence. In the proper use of the
word, the key term is discovery, commonly defined as "detecting something new". Often the more general terms "(large scale) data analysis" or "analytics" are more appropriate.

The actual data mining task is the automatic or semi-automatic analysis of large quantities of data in order to extract previously unknown interesting patterns such as groups of data records (cluster analysis), unusual records (anomaly detection) and dependencies (association rule mining). These patterns can then be seen as a kind of summary of the input data, and used in further analysis or for example in machine learning and predictive analytics. For example, the data mining step might identify multiple groups in the data, which can then be used to obtain more accurate prediction results by a decision support system. Neither the data collection, data preparation or result interpretation and reporting are part of the data mining step, but do belong to the overall data mining process as additional steps.

1.3 Association Rule Problem

Let $I = \{i_1, i_2, \ldots, i_m\}$ be a set of $m$ distinct attributes. Let $D$ be a database, where each record (tuple) $T$ has a unique identifier, and contains a set of items such that $T \subseteq I$. An association rule is an implication of the form of $X \Rightarrow Y$, where $X, Y \subseteq I$ are sets of items called itemsets, and $X \cap Y = \emptyset$. Here, $X$ is called antecedent while $Y$ is called consequent; the rule means $X \Rightarrow Y$. Association rules can be classified based on the type of values, dimensions of data, and levels of abstractions involved in the rule. If a rule concerns associations between the presence or absence of items, it is called Boolean association rule. And the dataset consisting of attributes which can assume only binary (0-absent, 1-present) values is called Boolean database.

2. EXPERIMENTAL RESULTS

For noninvariant similarities, the most well-known coefficient is the Jaccard dissimilarity coefficient, where the number of negative matches is considered unimportant and is ignored in the computation:

$$d(i, j) = \frac{r + s}{q + r + s}$$

To make experiments the data should be in binary format. The required dataset is obtained from Net. Two algorithms Apriori and SI were used to evaluate the data.

| Table 1: Transactional Database |
|-------------------|---|---|---|---|---|
| $T_1$ | $I_1$ | $I_2$ | $I_3$ | $I_4$ | $I_5$ |
| $T_2$ | 1 | 0 | 1 | 0 | 1 |
| $T_3$ | 1 | 1 | 0 | 1 | 1 |
| $T_4$ | 1 | 1 | 1 | 0 | 1 |

The algorithm was coded in ORACLE 10g using laptop computer having 20GB hard drive and 1.6MH processor. I create a table in the database to store the data for the purpose of experiment. To load the data to the database oracle provide a facility by making a control file and then by using sql loader. First the data was converted into a
format in which the item is separated by commas instead of spaces. Now the data is loaded to the table with the help of SQL loader.

0, 1, 0, 0, 1, 0, 0, 1  
1, 1, 0, 0, 1, 1, 0, 0  
0, 1, 1, 0, 0, 1, 0, 1  
1, 0, 0, 1, 0, 1, 1, 0

After loading the data into table the algorithms are implemented on the database having ten thousand records

Common steps from the Algorithms.

Input:

Φ User specified threshold (0 to 1)  
T Binary transactional Database.

Output:

<table>
<thead>
<tr>
<th>Min Support</th>
<th>Support count</th>
<th>Frequent item List</th>
</tr>
</thead>
<tbody>
<tr>
<td>20</td>
<td>4838</td>
<td>I1</td>
</tr>
<tr>
<td>20</td>
<td>3868</td>
<td>I2</td>
</tr>
<tr>
<td>20</td>
<td>5332</td>
<td>I3</td>
</tr>
<tr>
<td>20</td>
<td>4848</td>
<td>I4</td>
</tr>
<tr>
<td>20</td>
<td>4344</td>
<td>I5</td>
</tr>
<tr>
<td>20</td>
<td>1938</td>
<td>I1,I2</td>
</tr>
<tr>
<td>20</td>
<td>3878</td>
<td>I1,I3</td>
</tr>
<tr>
<td>20</td>
<td>3878</td>
<td>I1,I4</td>
</tr>
<tr>
<td>20</td>
<td>2423</td>
<td>I2,I3</td>
</tr>
<tr>
<td>20</td>
<td>1938</td>
<td>I2,I4</td>
</tr>
<tr>
<td>20</td>
<td>3393</td>
<td>I3,I4</td>
</tr>
<tr>
<td>20</td>
<td>2424</td>
<td>I4,I5</td>
</tr>
<tr>
<td>20</td>
<td>1938</td>
<td>I1,I2,I3</td>
</tr>
<tr>
<td>20</td>
<td>3393</td>
<td>I1,I3,I4</td>
</tr>
</tbody>
</table>

After performing the experiments the results produced were same. Changing threshold also has no effect over the output. So by applying these algorithms data from space science may be analyzed very critically and accurately.
Figure 2:

<table>
<thead>
<tr>
<th>Dissimilarity</th>
<th>Value</th>
<th>Frequent List</th>
</tr>
</thead>
<tbody>
<tr>
<td>.71</td>
<td>.80</td>
<td>I1,I2</td>
</tr>
<tr>
<td>.38</td>
<td>.80</td>
<td>I1,I3</td>
</tr>
<tr>
<td>.33</td>
<td>.80</td>
<td>I1,I4</td>
</tr>
<tr>
<td>.73</td>
<td>.80</td>
<td>I1,I5</td>
</tr>
<tr>
<td>.64</td>
<td>.80</td>
<td>I2,I3</td>
</tr>
<tr>
<td>71</td>
<td>.80</td>
<td>I2,I4</td>
</tr>
<tr>
<td>79</td>
<td>.80</td>
<td>I2,I5</td>
</tr>
<tr>
<td>.5</td>
<td>.80</td>
<td>I3,I4</td>
</tr>
<tr>
<td>.75</td>
<td>.80</td>
<td>I1,I2,I3</td>
</tr>
<tr>
<td>.53</td>
<td>.80</td>
<td>I1,I3,I4</td>
</tr>
</tbody>
</table>

3. CONCLUSION AND FUTURE WORK

By applying such techniques to space science data, reliable and accurate results were produced. Because the output in these techniques were based on mathematical threshold which is reliable and fast? In future such techniques may be applied to a huge amount of data. Such techniques may be used for the clustering measure in the said field. Discrimination in space may be minimized by this in term of its distribution and climate regarding issues.

4. REFERENCES

2. David Wai-Lok Cheung, Vincent T. Ng, Ada Wai-Chee Fu, and Yongjian, Fu (1996). Efficient Mining of Association Rules in Distributed Databases. IEEE Transactions on Knowledge and Data Engineering, 8(6), 911-922.
DATA QUALITY ENHANCEMENT USING DATA MINING APPROACH

Irshad Ullah  
Department of Computer Science, Lahore University of Management Science,  
Lahore, Pakistan  
Email: irshadullah79@gmail.com

ABSTRACT

Data Quality refers to the degree of excellence exhibited by the data in relation to the portrayal of the actual scenario. Data are deemed of high quality if they correctly signify the real-world construct to which they refer. Furthermore, apart from these definitions, as data volume increases, the question of internal consistency within data becomes paramount, regardless of fitness for use for any external purpose, e.g. a person's age and birth date may conflict within different parts of a database. The first views can often be in disagreement, even about the same set of data used for the same purpose. In this work, Data Mining algorithms will be used for the data quality enhancement process to prove experimentally and practically that how reliable, efficient and speedy are these for the enhancement of data quality? A solid mathematical threshold (0 to 1) is set to analyze the data. The obtained results will be tested by applying the approach to the databases and data warehouses of different sizes with different threshold values. The results produced will be of different magnitude from short to the largest sets of data items. By this, we may take the results produced for different purpose.

KEY WORDS

Data Quality; Experiment; Analysis; Excellence; Consistency.

1. INTRODUCTION

1.1 Data Analysis

Data analysis is a process in which raw data is prepared and structured so that valuable information can be extracted from it. The process of organizing and thinking about data is way to accepting what the data does and does not contain. There are a variety of ways in which public can approach data analysis, and it is notoriously easy to direct data during the analysis phase to push certain conclusions or agendas [7].

1.2 Data Mining

Generally, data mining (sometimes called data or knowledge discovery) is the process of analyzing data from different perspectives and summarizing it into useful information - information that can be used to increase revenue, cuts costs, or both. Data mining software is one of a number of analytical tools for analyzing data. It allows users to analyze data from many different dimensions or angles, categorize it, and summarize the relationships identified. Technically, data mining is the process of finding correlations or patterns among dozens of fields in large relational databases [10].
The Apriori algorithm [8] is a great achievement in the history of mining association rules. It is by far the most well-known association rule algorithm. This technique uses the property that any subset of a large item set must be a large item set.

The Off-line Candidate Determination (OCD) technique is proposed in [6], and is based on the idea that small samples are usually quite good for finding large itemsets. The OCD technique uses the results of the combinatorial analysis of the information obtained from previous passes to eliminate unnecessary candidate sets.

Sampling [5] reduces the number of database scans to one in the best case and two in the worst. A sample which can fit in the main memory is first drawn from the database. The set of large itemsets in the sample is then found from this sample by using a level-wise algorithm such as Apriori.

Each association rule mining algorithm assumes that the transactions are stored in some basic structure, usually a flat file or a TID list, whereas actual data stored in transaction databases is not in this form. All approaches are based on first finding the large itemsets. The Apriori algorithm appears to be the nucleus of all the association rule mining algorithms.

In this work my focus is on association rule mining technique. Two algorithms were selected, first the well known Apriori and our own developed SI [9] algorithm.

1.3 Logical data analysis
The logical analysis of data was originally developed for the analysis of datasets whose attributes take only binary (0-1) values [1, 2, 4]. Since it turned out later that most of the real-life applications include attributes taking real values, a “binarization” method was proposed in [3].

1.4 Binary variables
A binary variable has only two states: 0 or 1, where 0 means that the variable is absent, and 1 means that it is present. If all binary variables are thought of as having the same weight, we have the 2-by-2 contingency table of table 3, where w is the number of variables that equal 1 for both items i and j, x is the number of variables that equal 1 for item i but that are 0 for item j, y is the number of variables that equal 0 for item i but equal 1 for item j, and z is the number of variables that equal 0 for both item i and j. The total number of variables is V, where \( V = w + x + y + z \).

<table>
<thead>
<tr>
<th>Item i</th>
<th>Item j</th>
<th>Sum</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>w + x</td>
</tr>
<tr>
<td>0</td>
<td>w</td>
<td>x</td>
</tr>
<tr>
<td>Sum</td>
<td>y + z</td>
<td>V</td>
</tr>
</tbody>
</table>

For noninvariant similarities, the most well-known coefficient is the Jaccard dissimilarity coefficient, where the number of negative matches \( d \) is considered unimportant and thus is ignored in the computation:
\[ d(I, J) = \frac{b+c}{a+b+c} \quad (1.4.1) \]

The measurement value 1 suggests that the objects i and j are dissimilar and the measurement value 0 suggests that the objects are similar. This method is used in SI algorithm while the Apriori algorithm works using similarity measures.

2. EXPERIMENTAL RESULTS

Algorithms were coded in oracle 10g using laptop computer having 20GB hard drive and 1.6MH processor. A table was created in the database to store the data for the purpose of experiment. To load the data to the database oracle provide a facility by making a control file and then by using sql loader.

2.1 Common steps from the Algorithms.
Input:
\[ \Phi \text{ User specified threshold (0 to 1)} \]
\[ T \text{ Binary transactional Database.} \]

Output:
Frequent Items list.

The experiments were performed with two algorithms Apriori and SI. The results produced on different input threshold and database size given below.

**Figure 1:**

![Apriori results](image1)

**Figure 2:**

![SI results](image2)
The results produced by the algorithms were same. Changing size and input threshold do not affect the output. As the required data to be in binary format.

So the results produced are reliable and more efficient. Because 0 mean one state and 1 mean the other.

3. CONCLUSION AND FUTURE WORK

Results produced by the algorithm were same on different datasets. The data was in binary format. The output produced very fast and correct. In future more experiments may be performed for more enhancements.

4. REFERENCES

TESTING OF INTERCEPT WHEN SLOPE IS UNDER SUSPICION

Budi Pratikno$^1$ and Shahjahan Khan$^2$
Department of Mathematics and Computing
Australian Centre for Sustainable Catchments
University of Southern Queensland Toowoomba, Queensland, Australia
Emails: $^1$pratikto@yahoo.com.au and $^2$khans@usq.edu.au

ABSTRACT

In simple regression analyses, the inference on the intercept depends on the knowledge of the slope. This paper studies the problem of testing the intercept of a simple regression model when slope is under suspicion. Depending on the situation the slope may be unknown or unspecified, known or specified, and uncertain if the suspected value is unsure. The three different scenarios on the slope lead to three different tests of the intercept. Here we define the unrestricted test (UT), restricted test (RT) and pre-test test (PTT) for the intercept parameter depending on the level of knowledge on the slope. The test statistics, their sampling distributions, and power functions of the tests are derived and compared when the error variance is assumed to be known.

Keywords and phrases: Linear regression, test of intercept, power function, normal and bivariate normal distributions.

2010 Mathematics Subject Classification: Primary 62H12, Secondary 62G05.

1 Introduction

The simple regression model describes the linear relationship between dependent variable ($y$) and independent variable ($x$). For an $n$ pairs of observations, $(x_i,y_i)$, for $i = 1, 2, \ldots, n$ the model is written as

$$ y_i = \beta_0 + \beta_1 x_i + e_i, \quad (1.1) $$

where $e_i$'s are assumed to be normally distributed with mean 0 and variance $\sigma^2$. $x_i$'s are known real values of the independent variable, and $\beta_0$ and $\beta_1$ are the unknown intercept and slope parameters respectively. We consider the problem of testing $H_0 : \beta_0 = \beta_{00}$ (a fixed value) when there is uncertain information available on the value of $\beta_1$.

Inferences about population parameter could be improved using non-sample prior information (NSPI) from trusted sources (cf Bancroft, 1944). Such information, which is usually provided by previous studies or expert knowledge or experience of the researchers, and is not related to the sample data. An appropriate statistical test on the value (expressed in

---

$^1$On leave from General Soedirman University (UNSOED), Indonesia
the form) null hypothesis will be useful to eliminate the uncertainty on this suspected information. Then the outcome of the preliminary testing on the uncertain NSPI is used in the hypothesis testing to improve the performance of the statistical test (Khan and Saleh, 2001; Saleh, 2006; Yunus and Khan, 2010).

The suspected value of the slope may be (i) unknown or unspecified if NSPI is not available, (ii) known or specified if the exact value is available from NSPI, and (iii) uncertain if the suspected value is unsure. For the three different scenarios, three different of statistical tests, namely the (i) unrestricted test (UT), (ii) restricted test (RT) and (iii) pre-test test (PTT) are defined.

In the area of estimation with NSPI there has been a lot of work, notably Bancroft (1944, 1964), Hand and Bancroft (1968), and Judge and Bock (1978) introduced a preliminary test estimation of parameters to estimate the parameters of a model with uncertain prior information. Khan (2003, 2008), Khan and Saleh (1997, 2001, 2005, 2008), Khan et al (2002), Khan and Hoque (2003), Saleh (2006) and Yunus (2010) covered various work in the area of improved estimation using NSPI, but there is a very limited number of studies on the testing of parameters in the presence of uncertain NSPI. Although Tamura (1965), Saleh and Sen (1978, 1982), Yunus and Khan (2007, 2010), and Yunus (2010) used the NSPI for testing hypothesis using nonparametric methods, the problem has not been addressed in the parametric context.

This paper considers statistical tests with NSPI and the criteria that are used to compare the performance of the UT, RT and PTT are the size and power of the tests. A statistical test that has a minimum size is preferable because it will give a smaller probability of a Type I error. Furthermore, a test that has maximum power is preferred over any other tests because it guarantees the highest probability of rejecting any false null hypothesis. A test that minimizes the size and maximizes the power is preferred over any other tests. In reality, the size of a test is (kept) fixed, and then the choice of the best test is based on its maximum power.

We define the following three different tests:

For the UT, let $\phi^{UT}$ be the test function and $T^{UT}$ be the test statistic for testing $H_0 : \beta_0 = \beta_{00}$ (known constant) against $H_a : \beta_0 > \beta_{00}$ when $\beta_1$ is unspecified.

For the RT, let $\phi^{RT}$ be the test function and $T^{RT}$ be the test statistic for testing $H_0 : \beta_0 = \beta_{00}$ against $H_a : \beta_0 > \beta_{00}$ when $\beta_1$ is $\beta_{10} = 0$ (specified).

For the PTT, let $\phi^{PTT}$ be the test function and $T^{PTT}$ be the test statistic for testing $H_0 : \beta_0 = \beta_{00}$ against $H_a : \beta_0 > \beta_{00}$ following a pre-test (PT) on the slope. For the PT, let $\phi^{PT}$ be the test function for testing $H_{0}^{*} : \beta_1 = \beta_{10}$ (a suspected constant) against $H_{0}^{*} : \beta_1 > \beta_{10}$. If the $H_{0}^{*}$ is rejected in the PT, then the UT is used to test the intercept, otherwise the RT is used to test $H_0$. Thus, the PTT depends on the PT which is a choice between the UT and RT.

For the above tests we define following unrestricted estimator of $\beta_1$ and intercept $\beta_0$:

$$\widehat{\beta_1} = \frac{\sum_{i=1}^{n} (x_i - \overline{x})(y_i - \overline{y})}{\sum_{i=1}^{n} (x_i - \overline{x})^2} = \frac{s_{yx}}{s_{xx}}$$ and $$\widehat{\beta_0} = \overline{y} - \widehat{\beta_1}\overline{x},$$ where $\overline{x} = \frac{1}{n} \sum x$ and $\overline{y} = \frac{1}{n} \sum y$.

The restricted estimator (under $H_0^{*}$) of the slope and intercept are $\widehat{\beta_1} = \beta_{10}$ and $\widehat{\beta_0} =$
\[ Y - \hat{\beta}_1 \hat{X}. \]

The following section provides the three tests. Section 3 derives the distribution of the test statistics. The power functions of the tests are obtained in Section 4. An illustrative example is given in Section 5. The comparison of the power of the tests and concluding remarks are provided in Sections 6 and 7.

## 2 The Three Tests

For testing the intercept parameter under three different scenarios of the slope, the test statistics of the UT, RT and PTT for known \( \sigma^2 \) are given as follows. The test statistic of the UT for testing \( H_0 : \beta_0 = \beta_{00} \) against \( H_a : \beta_0 > \beta_{00} \) is defined by

\[
T^\text{UT}_z = \frac{\sqrt{n} \left( \bar{\beta}_0 - \beta_{00} \right)}{SE(\hat{\beta}_0)} = \frac{\sqrt{n} \left( \bar{y} - \bar{x} \hat{\beta}_0 \right)}{\sigma \left(1 + \frac{\bar{x}^2}{\sum x^2}\right)^{1/2}} = \frac{\sqrt{n} \left( \bar{y} - \bar{x} \beta_{00} \right)}{\sigma \left(1 + \frac{\bar{x}^2}{\sum x^2}\right)^{1/2}}, \tag{2.1}
\]

where standard error (SE) of \( \bar{\beta}_0 \) is \( \frac{\sigma}{\sqrt{n}} \left(1 + \frac{\bar{x}^2}{\sum x^2}\right)^{1/2} \). Under \( H_0 \), \( T^\text{UT}_z \) follows standard normal distribution \( N(0,1) \), and under \( H_a \) the distribution is \( N \left( \frac{\sqrt{n} \left( \beta_0 - \beta_{00} \right)}{\sigma \left(1 + \frac{\bar{x}^2}{\sum x^2}\right)^{1/2}}, 1 \right) \) with \( \beta_0 - \beta_{00} > 0 \) and \( \beta_{00} \) is the value of \( \beta_0 \).

The test statistic of the RT is given by

\[
T^\text{RT}_z = \frac{\left( \hat{\beta}_0 - \beta_{00} \right)}{SE(\hat{\beta}_0)} = \frac{\hat{\beta}_0 - \beta_{00}}{\sigma / \sqrt{n}} = \frac{\sqrt{n} \left( \bar{y} - \beta_{00} \bar{x} \right)}{\sigma \sqrt{n}} \sim N(0,1). \tag{2.2}
\]

Note that \( \hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} = \bar{y} - \beta_{10} \bar{x} \) and \( SE(\hat{\beta}_0) = \sqrt{Var(\hat{\beta}_0)} = \frac{\sigma}{\sqrt{n}} \). Under \( H_a \) the \( T^\text{RT}_z \) follows a normal distribution with mean \( \frac{\beta_0 - \beta_{00}}{\sigma / \sqrt{n}} \) and variance 1. If \( \beta_1 \) is specified to be \( \beta_{10} \) then under \( H_a \) it follows a normal distribution with mean \( \frac{(\beta_0 - \beta_{00}) + (\beta_1 - \beta_{10}) \bar{x}}{\sigma / \sqrt{n}} \) and variance 1, where \( \beta_0 - \beta_{00} > 0 \) and \( \beta_1 - \beta_{10} > 0 \).

For the preliminary test (PT) \( H^\ast_0 : \beta_1 = \beta_{10} \) against \( H^\ast_a : \beta_1 > \beta_{10} \), the test statistic of the PT is given by

\[
T^\text{PT}_z = \frac{\bar{\beta}_1 - \beta_{10}}{SE(\hat{\beta}_1)} = \frac{\bar{\beta}_1 - \beta_{10}}{\sigma / \sqrt{\sum x^2}} \sim N \left( \frac{\beta_1 - \beta_{10}}{\sigma / \sqrt{\sum x^2}}, 1 \right), \tag{2.3}
\]

where \( SE(\hat{\beta}_1) = \sigma / \sqrt{\sum x^2} \). Under the null hypothesis the above test statistic follows the standard normal distribution.

Furthermore, we propose the PTT for testing \( H_0 \), following the PT on \( \beta_1 \). Let us choose positive number \( \alpha_j \) \( (0 < \alpha_j < 1) \) and real values \( (z_{\alpha_j}) \), such that \( P \left( T^\text{PT}_z > z_{\alpha_1} \mid \beta_0 = \beta_{00} \right) = \alpha_1 \), \( P \left( T^\text{PT}_z > z_{\alpha_2} \mid \beta_1 = \beta_{00} \right) = \alpha_2 \) and \( P \left( T^\text{PT}_z > z_{\alpha_3} \mid \beta_1 = \beta_{10} \right) = \alpha_3 \). Then, the PTT for testing \( H_0 \) when \( \beta_1 \) is uncertain, it is given by the test function

\[
\Phi_z = \begin{cases} 
1, & \text{if } T^\text{PT}_z \leq z_{\alpha_2}, T^\text{RT}_z > z_{\alpha_2} \text{ or } T^\text{UT}_z > z_{\alpha_1} \\
0, & \text{otherwise.}
\end{cases} \tag{2.4}
\]
The size of the PTT is then
\[ \alpha_z = P \{ T_z^{PT} \leq z_{\alpha_3}, T_z^{RT} > z_{\alpha_1} \} + \{ T_z^{PT} > z_{\alpha_3}, T_z^{UT} > z_{\alpha_1} \} \] (2.5)

3 Distribution of Test Statistics

For the derivation of the power function of the UT and RT we need the sampling distributions of the $T_z^{UT}$ and $T_z^{PT}$, and that of the PTT the joint distribution of $(T_z^{UT}, T_z^{PT})$ and $(T_z^{RT}, T_z^{PT})$ are essential. Let \{ $K_n$ \} be a sequence of alternative hypotheses defined as
\[ K_n : (\beta_0 - \beta_{00}, \beta_1 - \beta_{10}) = \left( \frac{\lambda_1}{\sqrt{n}}, \frac{\lambda_2}{\sqrt{n}} \right) = n^{-1/2} \lambda, \] (3.1)
where $\lambda = (\lambda_1, \lambda_2)$ are fixed real numbers, $\beta_0$ is true value and $\beta_{00}$ is NSPI. Under $K_n$, the value of $\beta_0 - \beta_{00}$ is greater than zero (or $\beta_0 - \beta_{00} > 0$), and under $H_0$ the value of $\beta_0 - \beta_{00} = 0$.

Following Yunus and Khan (2011), the test statistic of the UT, under $K_n$, is $T_z^{UT} \sim N \left( \frac{\sqrt{n}(\beta_0 - \beta_{00})}{\sigma(1 + \frac{S_x^2}{S_z^2})^{1/2}}, 1 \right)$. Under alternative hypothesis we then derive $Z_i, i = 1, 2, 3$ as follows,
\[ Z_1 = T_z^{UT} - \frac{\sqrt{n}(\beta_0 - \beta_{00})}{\sigma(1 + \frac{S_x^2}{S_z^2})^{1/2}} = T_z^{UT} - \frac{\lambda_1}{k_1} \sim N(0, 1), \] (3.2)
where $k_1 = \sigma(1 + \frac{S_x^2}{S_z^2})^{1/2}$. Similarly, from equation (2.2) and (2.3), under $K_n$ we obtain
\[ Z_2 = T_z^{PT} - \frac{(\beta_0 - \beta_{00}) + (\beta_1 - \beta_{10}) \xi}{\sigma/\sqrt{n}} \sim N(0, 1), \] (3.3)
and
\[ Z_3 = T_z^{PT} - \frac{\beta_1 - \beta_{10}}{\sigma/\sqrt{S_x}} \sim N(0, 1). \] (3.4)

Since $T_z^{RT}$ and $T_z^{PT}$ are independent, the joint distribution under $H_\alpha$ is

\[
\begin{pmatrix}
T_z^{RT} \\
T_z^{PT}
\end{pmatrix}
\sim N_2 \left[ \begin{pmatrix}
\frac{(\beta_0 - \beta_{00}) + (\beta_1 - \beta_{10}) \xi}{\sigma/\sqrt{n}} \\
\frac{\lambda_1 + \lambda_2 \xi}{\sigma/\sqrt{n}}
\end{pmatrix}, \begin{pmatrix}
1 & 0 \\
0 & 1
\end{pmatrix} \right]
\]

\[ = N_2 \left[ \begin{pmatrix}
\frac{\lambda_1 + \lambda_2 \xi}{\sigma/\sqrt{n}} \\
\frac{\lambda_1 + \lambda_2 \xi}{\sigma/\sqrt{n}}
\end{pmatrix}, \begin{pmatrix}
1 & 0 \\
0 & 1
\end{pmatrix} \right]. \] (3.5)

In the same manner, we have
\[
\begin{pmatrix}
T_z^{UT} \\
T_z^{PT}
\end{pmatrix}
\sim N_2 \left[ \begin{pmatrix}
\frac{\sqrt{n}(\beta_0 - \beta_{00})}{\sigma(1 + \frac{S_x^2}{S_z^2})^{1/2}} \\
\frac{\lambda_1 + \lambda_2 \xi}{\sigma/\sqrt{n}}
\end{pmatrix}, \begin{pmatrix}
1 & -\rho \\
-\rho & 1
\end{pmatrix} \right]
\]
\[ = N_2 \left( \frac{\lambda_1}{\sigma(1 + \frac{\beta_1}{2\sigma})^{1/2}} \right), \left( \begin{array}{cc} 1 & -\rho \\ -\rho & 1 \end{array} \right) \right], \]  

(3.6)

where \( \rho \) is correlation coefficient between \( T_z^{UT} \) and \( T_z^{PT} \).

![Power of the tests](Image)

**Figure 1:** The power of the UT, RT and PTT against \( \lambda_1 \) with \( \beta_1 > 0 \) and \( \lambda_2 = 0, 1, 1.5, 2 \).
4 Power Functions and Size of Tests

The power function for the UT, RT and PTT for known variance are derived as below. The power function of the UT is

\[
P(\lambda) = \prod_{z_1} P(T_z^{UT} > z_{\alpha_1} | K_n) = 1 - G \left( z_{\alpha_1} - \frac{\lambda_1}{k_1} \right),
\]

and that of the RT is

\[
P(\lambda) = \prod_{z_2} P(T_z^{RT} > z_{\alpha_2} | K_n) = 1 - G \left( z_{\alpha_2} - \frac{\lambda_1 + \lambda_2}{\sigma} \right).
\]

When \(\lambda_1\) grows larger the power of the UT becomes higher. The power grows higher as \(\lambda_1\) becomes larger. Similarly, the power function of the PTT is given as

\[
P(\lambda) = \prod_{z} P(\text{rejecting } H_0) = \prod_{z} P(T_z^{PT} > z_{\alpha_3}) + P(T_z^{PT} > z_{\alpha_3} | T_z^{RT} > z_{\alpha_2})
\]

\[
= G \left( z_{\alpha_3} - \frac{\lambda_2 \sqrt{S_{xx}}}{\sigma \sqrt{n}} \right) \left( 1 - G \left( z_{\alpha_2} - \frac{\lambda_1 + \lambda_2}{\sigma} \right) \right) + d_{1,\rho} \left( z_{\alpha_3} - \frac{\lambda_2 \sqrt{S_{xx}}}{\sigma \sqrt{n}}, z_{\alpha_1} - \frac{\lambda_1}{k_1}; \rho \neq 0 \right),
\]

where \(d_{1,\rho}\) are bivariate normal probability integral. Here \(d_{1,\rho}\) is defined for every real \(p, q\) and \(-1 < \rho < 1\) as

\[
d_{1,\rho}(p, q, \rho) = \frac{1}{2\pi \sqrt{1 - \rho^2}} \int_0^\infty \int_0^\infty \exp \left( -\frac{1}{2(1-\rho^2)} (s^2 + y^2 - 2\rho s y) \right) dy ds,
\]

where \(p = z_{\alpha_3} - \frac{\lambda_2 \sqrt{S_{xx}}}{\sigma \sqrt{n}}, q = z_{\alpha_1} - \frac{\lambda_1}{k_1}\) and \(G(x)\) is a cumulative distribution function (cdf) of the standard normal distribution.

Furthermore, the size of the UT, RT and PTT are given as

\[
\alpha_z^{UT} = P(T_z^{UT} > z_{\alpha_1} | H_0) = 1 - G \left( z_{\alpha_1} - \frac{\sqrt{n}(\beta_0 - \beta_{00})}{\sigma \sqrt{[1 + n\bar{y}^2/S_{xx}]}}, H_0 : \beta_0 = \beta_{00} \right)
\]

\[
= 1 - G \left( z_{\alpha_1} - \frac{\sqrt{n}(\beta_0 - \beta_{00})}{k_1} \right) = 1 - G \left( z_{\alpha_1} \right).
\]
Figure 2: The size of the UT, RT and PTT against $\lambda_1$ with $\lambda_2 = 0, 1, 1.5, 2$.

\[
\alpha_{\alpha}^{RT} = P(T_{\alpha}^{RT} > z_{\alpha} \mid H_0) = 1 - G\left(z_{\alpha} - \frac{\lambda_2 \bar{X}}{\sigma}\right), \quad \text{and} \quad (4.6)
\]

\[
\alpha_{\alpha}^{PTT} = G(z_{\alpha})\left(1 - G\left(z_{\alpha} - \frac{\lambda_2 \bar{X}}{\sigma}\right)\right) + d_{1,\rho}(z_{\alpha_3}, z_{\alpha_1}, \rho \neq 0). \quad (4.7)
\]
5 A Simulation Example

To study the properties of the three tests we conduct a simulation study. The main aim is to compute the power function of the tests and compare them graphically. In this simulated example we generate random data using R package. The independent variable \(x\) and error \(e\) are generated from the uniform distribution with \(a = 0\), and \(b = 1\) and from normal distribution with \(\mu = 0\) and \(\sigma^2 = 1\), respectively. In each case \(n = 20\) random variates were generated. The dependent variable \(y\) is then determined by \(y = \beta_0 + \beta_1 x + e\) for \(\beta_0 = 5\) and \(\beta_1 = \pm 2.5\). For the computation of the power functions of the tests we set \(\alpha_1 = \alpha_2 = \alpha_3 = \alpha = 0.05\). The graphs for the power functions and size of the three tests for known error variance are provided by using the formulas in (4.1), (4.2), (4.3), (4.5), (4.6) and (4.7). Identical graphs for the power and size curves are observed when the slope is negative.

6 Comparison of Power and Size

From Figure 1, as well as from equation (4.1) we see that the power of the UT does not depend on \(\lambda_2\) and \(\rho\) but it increases as the value of \(\lambda_1\) increases. Its form is sigmoid, starting from a very small value of near zero at \(\lambda_1 = 0\), it approaches 1 when \(\lambda_1\) is large (about 20 in Figure 1 and 2). Thus the power of the UT changed significantly for any value of \(\lambda_1\) from 0 to 20. The minimum power of the UT is around 0.05 for \(\lambda_1 = 0\). The power curve of the RT is also sigmoid for all values of \(\lambda_1\) and \(\lambda_2\). The power of the RT increases as the values of \(\lambda_1\) and/or \(\lambda_2\) increase. Moreover, the power of the RT is always larger than that of the UT and PTT for all values of \(\lambda_1\) and/or \(\lambda_2\). The minimum power of the RT is around 0.05 for \(\lambda_2 = 0\) (as well as for \(\lambda_1 = 0\)) and increases to be around 0.1 for \(\lambda_2 = 2\). The maximum power the RT is around 1 for \(\lambda_1\) around 10 or above. The power of the PTT also depends on the values of \(\lambda_1\) and \(\lambda_2\). Like the power of the RT, the power of the PTT increases for large value of \(\lambda_1\) and tend to decrease as \(\lambda_2\) grows larger. Moreover, the power of the PTT tend to be larger than that of the UT. The minimum power of the PTT is around 0.05 for \(\lambda_2 = 0\) and \(\lambda_1 = 0\), and it increases to be around 0.1 for \(\lambda_2 = 2\). The gap between the power curves of the RT and PTT is obviously clear for all values of \(\lambda_1\) and \(\lambda_2\). Like the power of RT, the power of PTT depends on any values of \(\lambda_1\) and \(\lambda_2\).

Figure 2 or equation (4.5) shows the size of the UT does not depend on \(\lambda_2\). It is constant and remains unchanged for all values of \(\lambda_1\) and \(\lambda_2\). The size of the RT is also constant for all values of \(\lambda_1\). However, the size of the RT increases as the value of \(\lambda_2\) increases. Moreover, the size of the RT is always larger than that of the UT for all values of \(\lambda_2\) except for \(\lambda_2 = 0\) when both tests have the same size. Like the size of the RT, the size of the PTT increases as \(\lambda_2\) grows larger. The difference between the size of the RT and PTT does not change much as the value of \(\lambda_2\) increases. The size of the RT and PTT increase as the value of \(\lambda_2\) increases. Also, the size of the RT is larger than that of the UT and PTT except for \(\lambda_2 = 0\), when they are the same.
Figure 3 shows the power of the PTT and size against $\rho$ and $\lambda_2$. It increases significantly as the value of $\rho$ increases from $\rho = -1$ to $\rho = 0$ and stays the same for $\rho$ from zero to $1$. The difference of the power of the PTT for $\lambda_2 = 0$ and $\lambda_2 = 2$ is significantly different as the value of $\rho$ increases. From this figure or equation (4.6 and 4.7) the size of the RT and PTT depend on the value of $\lambda_2$. They increase as the value of $\lambda_2$ increases. Unlike the size of the RT and PTT, the size of the UT does not depend on the value of $\lambda_2$ and it remains constants for all the values of $\lambda_2$. The size of the RT is always greater than the size of the
Testing of intercept when slope is under suspicion

7 Concluding Remarks

Based on the above analyses, the power of the RT is always higher than that of the UT and PTT for all values of $\lambda_1$, and the power of the PTT lies between the power of the RT and UT for all values of $\lambda_1$, $\lambda_2$, and $\rho$. The size of the UT is smaller than that of the RT and PTT. The RT has maximum power and size, and the UT has minimum power and size. The PTT has smaller size than the RT and the RT has larger power than the UT. The PTT protects against maximum size of the RT and minimum power of the UT.

As $\lambda_2 \to 0$ the difference between the power of the PTT and RT diminishes for all values of $\lambda_1$. That is, if the NSPI is accurate the power of the PTT is about the same as that of the RT. Moreover, the power of the PTT gets closer to that of the RT as $\rho \to 1$.

The size of the PTT becomes smaller as $\lambda_2 \to 0$. Once again if the NSPI is near accurate the size of the PTT approaches that of the UT. Therefore, we recommend PTT when the quality of the NSPI is good (i.e. $\lambda_2 \to 0$) and it performs even better than the UT and RT when $\rho \to 1$.

REFERENCES


TRENDS IN RESEARCH PUBLICATIONS OF PAKISTANI STATISTICAL JOURNALS

Mehwish Hussain and Waqas Ahmed Farooqui

1 Department of Statistics, University of Karachi, Karachi, Pakistan.
Email: mehvish.hussain@yahoo.com
2 Department of Research, Dow University of Health Sciences, Karachi, Pakistan. Email: waqasahmed.farooqui@gmail.com

ABSTRACT

The academic growth of any country is measured by the research works published by her indigenous authors. The quality of research publications is usually measured by impact factor which itself depends on number of citations for corresponding journal articles. The impact factor of Pakistani journals usually remains low hitherto. One of the stated reasons is worldwide unfamiliarity of research outputs from this region. Many articles published in international journals stipulated the trends of research publication in statistical journals of their regions. This research study aims to scrutinize the trend of research publications by three Pakistani statistical journals; Pakistan Journal of Statistics (PJS), Journal of Statistics (JS) and Pakistan Journal of Statistics and Operation Research (PJSOR). The online available information noted down which are year of publication, month of publication, volume number, issue number, total number of publications in each issue & year, title of the study, area of statistics discussed and authors & co-authors’ affiliations with respect to their country, continent and country religion belonging (Islamic/Non-Islamic). Trends in publications are found out and hierarchical models of research outputs are constructed which are stratified by the three journals, author’s affiliation and topic of statistics.

KEYWORDS

Trend; statistical journal; publication; Pakistan; impact factor.

INTRODUCTION

Worldwide, a dramatically growth is observed from late 90’s in research work in world of science (Link and Scott, 2003). The academic growth is now being measured by research outputs produced by any institute or country. Indeed, one of the measures of university ranking is research publications made by the university personnel (HEC-PAK, 2011, ARWU, 2011). There are many measures of quality of research works published in scientific journals. Impact factor and h-indices are the instances for the same. These indices depend on number of citation made for an article of the corresponding journal (Mahbuba and Rousseau, 2010, Van Nierop, 2009). The criteria of citation of an article are usually based on different features. Some of them are: reputation of journals in which the related article appears, the familiarity of author(s), the institutional affiliation of the
authors, the type of publisher of the article, the online accessibility of the article, the source of the journal (refereed journals or non-journal sources) (Tenopir et al., 2010).

Statistical journals have usually low impact factor. Even the top statistical journal, Journal of the Royal Statistical Society Series B – Statistical Methodology, ranked at place 1046 with impact factor equaled 3.50 comparing with the CA: A Cancer Journal for Clinician having highest impact factor equaled 96.242 for the year 2010 (JCR, 2010). The reported reason for this was articles in statistical journals take longer time to be cited than other disciplined journals. It may be due to interlink of the topics published in medical journals (Van Nierop, 2009).

Pakistani journals have notably ranked in lower category. Only six Pakistani journals were assigned impact factors with less than 1.0 unit in JCR, 2010. Pakistan Journal of Statistics (PJS) was the only statistical journal ranked 5th among them with an impact factor of 0.156. Reasons could be derived with the earlier analyses i.e. presence of large number of un-cited articles from Pakistani journals or less participation in international collaboration (Mahbuba and Rousseau, 2010). Another undocumented reason can be unfamiliarity of research outputs from Pakistan as most of the Pakistani journals are not indexed in reputed search engines such as Google Scholar, Web of Science, Directory of Open Access Journals (DOAJ), Springer, Elsevier, Science Direct etc. One other probable reason is the online unavailability of research output (in form of article, journal or webpage) from Pakistan.

Gil\textsuperscript{1}, Genest and Baltagi updated the trend in statistical publications with respect to geography or field of statistics (Gil et al., 1999, Gil et al., 2000, Genest and Guay, 2002, Baltagi, 2007). No study has been done on exploring the statistical research output publishing in countries like Pakistan.

Keeping these aspects in mind, this study was designed to elaborate the research output published in Pakistani journals devoted to the subject Statistics. This study focused on trend in the publication in three statistical journals of Pakistan, mainly stratifying the author’s affiliations and topic of statistics in these publications. Trends in publications among different statistical fields were found out. It was derived to enhance the knowledge of collaborative work of Pakistani authors with other countries, regions and with respect to topic of statistics.

**METHODOLOGY**

The online available data were noted down from the websites of three Pakistani Statistical Journals: Pakistan Journal of Statistics (PJS), Journal of Statistics (JS) and Pakistan Journal of Statistics and Operation Research (PJSOR). A brief description of these journals is given below:

- Pakistan Journal of Statistics (PJS), published quarterly under Islamic Countries Society of Statistical Sciences (ISOSS) since 1985.

\textsuperscript{1} Gil (1999\textsuperscript{a}, 2000\textsuperscript{b})

The information noted down were: year of issue, volume number, issue number, issuing month, total number of publications in each issue & year, title of the study, area of statistics discussed and authors & co-authors’ affiliations with respect to their country, continent and country religion belonging (Islamic/Non-Islamic).

A description on allotting area of statistics used in research and affiliations of authors are described below:

The area of statistics was allotted using classification in Wikipedia. Some of the areas were also included such as biostatistics and applied statistics. These areas were derived while scrutinizing the title of the study. If any study was covering more than one area, then classification were made on the basis of major and minor area covered. Though, in this article only major areas were presented.

Author’s affiliation was notified if it is already mentioned in the selected article of the journal. Alternatively, affiliations of authors were made while browsing their names in different online search engines such as Academic Research Microsoft, Google Scholar and Main Google webpage. The author’s affiliation was set unknown if it was not acquired after browsing these websites (Google, 2011, Microsoft, 2011, Scholar, 2011).

STATISTICAL ANALYSIS

Data was collected and compiled in Microsoft Excel v. 2010. Descriptive statistics and graphical presentation was also done in the same software. Analytical part of the research was run in Predictive Analytical Software (PASW 18.0). Hierarchical Regression analysis was performed while setting number of publications as dependent variable and year of publication as independent variable. Journal, area of statistics, Pakistani affiliation and selected topic of statistics were set as explanatory variables.

RESULTS

A total of 850 articles were available online in the three Pakistani Statistical Journals websites. Nine hundred and twenty seven authors shared their research works in these journals.

PJS published highest number of articles i.e. 715 articles. On the other hand, 102 and 33 articles were published in PJSOR and JS\(^2\). Most of the articles were written on the topics covering the areas Probability Distribution and Estimation Theory. Mathematical Statistics, Frequentist Inference, Regression Analysis and Sampling Methodology were the topics standing next in the series of number of publications (Table 1).

Three hundred articles were single author study whilst 369 were written by two authors. Only 3 studies were published by with the collaborative work of maximum 5 authors. Around 29% (246) articles contained at least one Pakistani author. On the other hand, 675 (79.4%) articles were written by at least single non-Pakistani authors. About

\(^2\) The online available journals from Journal of Statistics were till the year 2009.
72.3% and 59.1% articles were written by authors belonging to Asia and Islamic countries respectively (Table 2).

The research publications were contributed by 51 countries by 927 authors. Among those 764 authors published their articles in PJS, 128 in PJSOR and 35 in JS. After Pakistan, majority of authors resided in United States of America (USA), India, Saudi Arabia, Iran and Canada. Very few articles were published by South America. Asia following North America and Europe contributed in majority.

The collaboration of USA resided authors was seen in all areas of statistics. Her collaboration was more in Estimation Theory. India produced least collaboration amongst other. Nonetheless, her collaboration was high in the field of probability distribution. Saudi Arabia also collaborated high in this field. Design of Experiment was also the field highly contributed by her (Fig. 1).

Trend in the publications in Pakistani journals showed a linear increase. The linear regression analysis revealed that with the increment of one year, 1.142 fold increased in number of publications in Pakistani statistical journals. PJS, being a leading statistical journal of Pakistan, had remarkably high publication in the years 2009 and 2010. Though, as it can be seen in figure (2), during 1998 to 2002, there were a slight fall in the publications of PJS. Overall average number of article publication in PJS was 31.48 ± 14.214. The regression coefficient for PJS publication was 0.603 per year. JS had around constant number of publications throughout the year. Around 6.60 ± 1.342 articles published each year in JS. The regression coefficient revealed a decrement of 0.60 yearly publications in JS. PJSOR initiated with noticeably good number of publications. In 2009, it had less publication. However, this trend increased upward afterward. Noticeably, it had more publication in 2011 than PJS. The average number of articles published in PJSOR each year is 14.57 ± 5.884. Nevertheless, regression analysis estimated 0.893 augmentations in yearly publication of PJSOR (Fig. 2).

Topic wise trend analysis stipulated that the researches in the field of probability distribution were highest throughout the years except in 1988 (β = 0.173). Estimation theory followed the same trend with zero number of publications in 1991 (β = 0.124). Mathematical Statistics was the field in which publications were initiated after 1990. In 1997, there was zero publication in this area (β = 0.305). The research output in the other fields such as Regression Analysis, Frequentist Inference and Sampling Methodology were also high. However, there were many ups and downs in the publications of these fields during different eras of times (Fig. 3).

**CONCLUSION**

There is a strikingly increase in the research publication in the three Pakistani Statistical Journals. Pakistani authors enhance their contributions with other countries to produce quality research in the field of Statistics. The journals and articles should be indexed online such that research outputs form Pakistan can be cited by others. There should be more collaborative work and quality researches should be published in these journals. Training to the junior researchers should be given to enhance contributions in the publication in Pakistani Statistical Journals. They should be given more chances to present their research works for the improvement. It will enhance the development in research publication of the subject “Statistics”.
### Table 1:
**Number of Article Published in three Pakistani Statistical Journals and Area of Statistics Covered**

<table>
<thead>
<tr>
<th>Journal†</th>
<th>PD</th>
<th>ET</th>
<th>MS</th>
<th>FI</th>
<th>RA</th>
<th>SM</th>
<th>DOE</th>
<th>Demo</th>
<th>TS</th>
<th>DM</th>
<th>BI</th>
<th>Bio</th>
<th>RV</th>
<th>CS</th>
<th>MA</th>
<th>LF</th>
<th>SG</th>
<th>NP</th>
<th>Other</th>
</tr>
</thead>
<tbody>
<tr>
<td>PJS</td>
<td>715</td>
<td>102</td>
<td>33</td>
<td>156</td>
<td>124</td>
<td>73</td>
<td>62</td>
<td>60</td>
<td>55</td>
<td>47</td>
<td>45</td>
<td>35</td>
<td>34</td>
<td>31</td>
<td>23</td>
<td>19</td>
<td>16</td>
<td>12</td>
<td>11</td>
</tr>
<tr>
<td>PJSOR</td>
<td>102</td>
<td>715</td>
<td>131</td>
<td>215</td>
<td>186</td>
<td>156</td>
<td>124</td>
<td>102</td>
<td>98</td>
<td>84</td>
<td>81</td>
<td>73</td>
<td>60</td>
<td>55</td>
<td>47</td>
<td>45</td>
<td>35</td>
<td>34</td>
<td>31</td>
</tr>
<tr>
<td>JS</td>
<td>33</td>
<td>33</td>
<td>125</td>
<td>156</td>
<td>186</td>
<td>124</td>
<td>73</td>
<td>62</td>
<td>60</td>
<td>47</td>
<td>45</td>
<td>35</td>
<td>34</td>
<td>31</td>
<td>23</td>
<td>19</td>
<td>16</td>
<td>12</td>
<td>11</td>
</tr>
</tbody>
</table>


### Table 2:
**Number of Article Publication based on Pakistani/Non-Pakistani, Asian/Non-Asian and Islamic/Non-Islamic Belongings**

<table>
<thead>
<tr>
<th>Number of Authors</th>
<th>Total</th>
<th>Pakistani</th>
<th>Non-Pakistani</th>
<th>Asian</th>
<th>Non-Asian</th>
<th>Islamic</th>
<th>Non-Islamic</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>f</td>
<td>%</td>
<td>f</td>
<td>%</td>
<td>f</td>
<td>%</td>
<td>f</td>
</tr>
<tr>
<td>0</td>
<td>-</td>
<td>-</td>
<td>604</td>
<td>175</td>
<td>235</td>
<td>454</td>
<td>347</td>
</tr>
<tr>
<td>1</td>
<td>300</td>
<td>35.29%</td>
<td>112</td>
<td>294</td>
<td>301</td>
<td>259</td>
<td>252</td>
</tr>
<tr>
<td>2</td>
<td>369</td>
<td>43.41%</td>
<td>93</td>
<td>268</td>
<td>219</td>
<td>97</td>
<td>180</td>
</tr>
<tr>
<td>3</td>
<td>147</td>
<td>17.29%</td>
<td>36</td>
<td>93</td>
<td>81</td>
<td>34</td>
<td>61</td>
</tr>
<tr>
<td>4</td>
<td>31</td>
<td>3.65%</td>
<td>5</td>
<td>17</td>
<td>14</td>
<td>5</td>
<td>10</td>
</tr>
<tr>
<td>5</td>
<td>3</td>
<td>0.35%</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>1</td>
<td>2</td>
</tr>
</tbody>
</table>
Fig. 1: Collaboration of Pakistani Authors with other Countries on Different Topics of Statistics

Fig. 2: Number of Publications in Three Pakistani Statistical Journals since their Inceptions
Fig. 3: Trend in Publications in Three Pakistani Statistical Journals in Different Area of Statistics
REFERENCES

INSTRUMENTAL VARIABLE ESTIMATOR OF THE SLOPE PARAMETER WHEN THE EXPLANATORY VARIABLE IS SUBJECT TO MEASUREMENT ERROR

Anwar Saqr and Shahjahan Khan
Department of Mathematics and Computing
Australian Centre for Sustainable Catchments
University of Southern Queensland Toowoomba
Australia.
Email: anwar.saqr@usq.edu.au and khans@usq.edu.au

ABSTRACT

This paper proposes a new instrumental variable to estimate the parameters of a simple linear regression model where the explanatory variable is subject to measurement error. The new instrumental variable is defined using reflection of the observed values of the explanatory variable. Like other instrumental variable estimators, it is unbiased and consistent, but over performs estimators proposed by Wald (1940), Bartlett (1949), and Durbin (1954) if the ratio of the error variances is equal or less than one $\lambda \leq 1$. The method is straightforward, easy to implement, and performs much better than the existing instrumental variable based estimators. The theoretical superiority of the proposed estimator over the existing instrumental variable based estimators is established by analytical results of simulation. Two illustrative examples for numerical comparisons of the results are also included.

Keywords and phrases: Regression parameters, Measurement error models, Instrumental variable, Reflection of point; Reliability ratio; Sum of square error.

2010 Mathematics Subject Classification: 62F10 and 62J05.

1 Introduction

The linear regression model is arguably the most frequently used statistical tool in various fields of scientific investigations, including bioassays and econometric studies. Commonly used bioassays where regression model can be used may include prediction of the body weight based on body fat, or the yield of a crop based on soil moisture level. However, measuring the explanatory variable, namely, the body fat or soil moisture level is likely to involve measurement errors. The ordinary least squares (OLS) estimator of the regression

1On leave from Department of Statistics, AlJabal AlGarby University, Charian, LIBYA
parameters is inappropriate (biased and inconsistent) in the presence of measurement error or error in variable. As a result, in real life, measurement error poses a serious problem, as it directly impacts on estimators of the parameters and their standard error. It is well known that the measurement error in the response variable is not as serious as it is in the explanatory variable. The errors in the response variable can be absorbed in the error term of the model. The error in the explanatory variable causes various problems, and requires to be handled appropriately.

The measurement error is a real problem and it has been considered by a host of authors since the third quarter of the nineteenth century. Adcock (1877, 1878) discussed the problem in the context of least squares method. Pearson (1901) suggested some estimators based on Adcock’s work. The problem has been seriously considered by researchers from the later part of the first half of the last century. Wald (1940), Bartlett (1949), Durbin (1954), and Riggs et al. (1978) considered fitting regression line when both variables are subject to error. Berkson (1950) noted that if there is error in the explanatory variable the bias in the estimated regression line will be there regardless of the data being a random sample or the population. Burr (1988) considered error in explanatory variable for the binary responses model. Freedman et al. (2004) suggested a reconstructed moment base method to deal with error in the explanatory variable. The problem of error in both explanatory and response variables was considered by Geary (1942), Madansky (1959) and Halperin (1961).

Instrumental variable (IV) technique is a well known method to obtain unbiased and consistent estimators in the presence of measurement error in the explanatory variable. The method requires to define an IV that is uncorrelated with the model error but highly correlated with the explanatory variable. The grouping method was first suggested by Wald (1940), followed by Bartlett (1949) and then Durbin (1954). Maddala (1988) showed that Wald method is equivalent to using the instrumental variable $Z$ equal -1 and +1 for values less than or greater than the median of the manifest variable. Bartlett proposed to divide the values in three equal groups and use the first and third groups, and Durbin used the ranks of the values to define the IV. In each of the method there is loss of information (for not using actual values and dropping some of the data points), and there are different formulae to find the sum of squares error, and hence lead to different mean sum of square error, making the analysis incomparable.

In this paper we propose a new way to define IV using the reflection of the explanatory variable. The estimator based on this method is unbiased and consistent. Moreover, it allows to define the sum of squares error uniquely, same way as in the case of no measurement error. In addition there is no loss of information in this method. Both analytical results and numerical illustrations confirms the superiority of the proposed (instrumental variable) estimator over the existing estimators.

Degracie and Fuller (1972) considered estimation of the slope and covariance when the concomitant variable is measured with error. Grubbs (1973) discussed errors of measurement, precision, accuracy and the statistical inference. Aigner (1973) considered regression with a binary variable subject to errors of observation. Florens et al. (1974) considered

The estimation methods suggested by the above studies make assumption that the variances of the explanatory variable without and with measurement errors or the reliability ratio of the two variances are known (cf Fuller 2006, p.5). An alternative assumption is that the variance of the measurement error is known. All these assumptions are unrealistic and the methods based on them are not free from the restrictions imposed by the assumptions.

In this paper we propose a new instrumental variable method to estimate the parameters of the simple regression model without making any of the above assumptions on the variances of the explanatory variable. The proposed method uses the reflection of the manifest values of the explanatory variable. The reflection points about the regression line are defined by using transformation formula involving sin and cos functions. The use of the reflections of the observed values of the explanatory variable in defining IV provide a much better estimator of the slope and intercept parameters. It also reduces the mean sum of squares error. The analysis of variance and regression inferences based on the reflections have much better statistical properties than that using the observed values of the explanatory variable, or any other IV estimator.

In the next Section the measurement error regression model is introduced. Section 3 covers the existing estimation methods for the measurement error model. The proposed new estimator based on the reflections of the observed values of the explanatory variable is provided in Section 4. The superior properties of the new estimators are discussed in Section 5. Two numerical illustrations are provided in Section 6, and some concluding remarks are included in Section 7.

2 Measurement error models

In the conventional notation, let $X$ denote the true measurement on the explanatory variable. This is also called the latent variable. In the presence of measurement error the actual
observations are different from $X$. Let $M$ be the observable, or manifest variable of the explanatory variable. When the true value of the latent variable $X$ is observed, the commonly used classical simple linear regression model is represented by

$$Y_j = \beta_0 x + \beta_1 X_j + e_j, \quad j = 1, 2, \ldots, n,$$

where $Y_j$ is the $j$th realization of the response variable, $X_j$ is the fixed $j$th value of the explanatory variable, and $e_j$ is the error variable for $j = 1, 2, \ldots, n$. It is assumed that the model errors are independently distributed according to normal law with constant but unknown variance, that is, $e_j \sim N(0, \sigma_{e2})$.

If there is error in the explanatory variable, the actual observed value, $M$, is not the ‘true’ value of the explanatory variable. When the observed value of the explanatory variable contains measurement error, we define

$$M_j = X_j + u_j, \quad j = 1, 2, \ldots, n$$

where $u_j$ is the measurement error, and is assumed to be distributed as $N(0, \sigma_{uu})$. Note that, unlike $X_j$, $M_j$ is a random variable which is assumed to be distributed as $N(\mu_m, \sigma_{mm})$. The model with the fixed $X$ is called the functional model, and that with the random or stochastic $M$ is called the structural model.

The simple regression model with measurement error in the explanatory variable can be expressed as

$$Y_j = \beta_0 x + \beta_1 x M_j + v_j, \quad j = 1, 2, \ldots, n,$$

where $v_j = e_j - \beta_1 x u_j$. Note in equation (2.1) $X$ and $e$ are independent but in equation (2.3) $M$ and $v$ are not independent. So the application of least squares method is not valid for the models with measurement error. Thus, unlike for the model in (2.1), the validity of the estimator of the slope and intercept of the model in (2.3) is not obvious. However, Fuller (2006, p.3) assumes that $u_j$, $X_j$ and $e_j$ are mutually independent for the estimation of parameters. It also assumes that the reliability ratio, $\kappa_{xm} = \sigma_{mm}^{-1} \sigma_{xx}$, where $\sigma_{mm}$ is the variance of the manifest variable $M$, and $\sigma_{xx}$ is the variance of the latent variable $X$, is known.

## 3 Existing Estimators of parameters

The ordinary least squares (OLS) estimator of the regression parameters for the functional model are

$$\hat{\beta}_1 x = \frac{S_{xy}}{S_{xx}}, \quad \text{and} \quad \hat{\beta}_0 x = \bar{Y} - \hat{\beta}_1 x \bar{X},$$

where

$$S_{xy} = \sum_{j=1}^{n} (X_j - \bar{X})(Y_j - \bar{Y}), \quad S_{xx} = \sum_{j=1}^{n} (X_j - \bar{X})^2,$$
in which \( \hat{Y} = n^{-1} \sum_{j=1}^{n} Y_j \) and \( \hat{X} = n^{-1} \sum_{j=1}^{n} X_j \). The estimators of slope and intercept parameters are linear functions of the responses, and they are well known best linear unbiased estimators.

The sampling distribution of the estimator of the regression parameters is given by

\[
\begin{pmatrix}
\hat{\beta}_{0x} \\
\hat{\beta}_{1x}
\end{pmatrix}
\sim N_2 \left( \begin{pmatrix}
\beta_{0x} \\
\beta_{1x}
\end{pmatrix}, \sigma_{ee} \left( \begin{pmatrix}
\frac{1}{n} + \frac{\hat{X}^2}{S_{xx}} & \frac{-\hat{X}}{S_{xx}} \\
\frac{-\hat{X}}{S_{xx}} & \frac{1}{S_{xx}}
\end{pmatrix} \right) \right).
\]

(3.3)

The unbiased estimator of the error variance \( \sigma_{ee}^2 \) is given by \( \sigma_{ee}^2 = (n-2)^{-1}SSE_e = s_{ee}^2 \), where \( SSE_e = \sum_{j=1}^{n}(Y_j - \hat{Y}_j)^2 \) in which \( \hat{Y}_j = \hat{\beta}_{0x} + \hat{\beta}_{1x} X \) is the estimated value of \( Y_j \). Also, \( \sigma_{ee}^{-1}SSE_e \) follows a \( \chi^2 \) distribution with \((n-2)\) degrees of freedom.

In the presence of measurement error, the \( M \) values are observed instead of \( X \).

The least squares method yields the fitted model to be

\[
\hat{Y}_m = \hat{\beta}_{0m} + \hat{\beta}_{1m} M,
\]

(3.4)

where \( \hat{\beta}_{1m} = \frac{\hat{S}_{xy}}{\hat{S}_{mm}} = \hat{\beta}_{1x} k_{zm} \) in which \( k_{zm} = \frac{\hat{S}_{zm}}{\hat{S}_{mm}} \). It can be easily shown that \( \hat{\beta}_{1m} \) is a biased estimator of \( \beta_{1x} \). Also, the above estimator is not a consistent estimator of \( \beta_{1x} \). But there are other estimators in the literature that are unbiased and consistent. The instrumental variable (IV) method provides such an unbiased and consistent estimator.

Note that the regression parameters are different for the model with the manifest variable than that with the latent variable. Even though the aim is to estimate and test \( \beta_{0x} \) and \( \beta_{1x} \), but in reality one may end up estimating and testing \( \beta_{0m} \) and \( \beta_{1m} \) if we fully rely on \( M \), and over look the presence of the measurement error. In the literature, the regression parameters are estimated, for observed \( X \) values, under certain assumptions. One of the assumptions is that the reliability ratio, \( r_{mm} \) is known. Fuller (2006) used this assumption for the estimation of the regression parameters for the functional model.

### 3.1 Instrumental variable estimator

In the presence of measurement error in the explanatory variable the IV estimator for the regression parameters is defined as

\[
\hat{\beta} = (Z'M)^{-1}Z'y.
\]

(3.5)

where \( \hat{\beta} = (\hat{\beta}_0, \hat{\beta}_1)' \) is the vector of estimator of the intercept and slope parameters of the model \( M = \begin{pmatrix}
1 & 1 & \cdots & 1 \\
\bar{X}_1 & \bar{X}_2 & \cdots & \bar{X}_n
\end{pmatrix} \) and \( Z = \begin{pmatrix}
1 & 1 & \cdots & 1 \\
\bar{z}_1 & \bar{z}_2 & \cdots & \bar{z}_n
\end{pmatrix} \) in which \( \bar{z}_j \)'s are the values of the second row of the instrumental variable \( Z \). The selection of the values of \( \bar{z}_j \)'s require that it is highly correlated with the explanatory variable but uncorrelated with the model errors. The above IV estimator is unbiased and consistent. The variance of the above estimator is given by

\[
\text{Var}(\hat{\beta}) = \sigma_{ee}^2(Z'M)^{-1}(Z'Z)(Z'M)^{-1}.
\]

(3.6)
Obviously the value of the estimator and the variance depend on the choice of $Z$. For instance, Wald method, as suggested by Maddala (1988), defines $Z$ by assigning $z_j$ to be -1 or +1 depending on $m_j$ being smaller or larger than the median value of the manifest variable. The estimator of slope under this choice of IV is \( \hat{\beta}_{1W} = (\bar{Y}_2 - \bar{Y}_1)/(\bar{M}_2 - \bar{M}_1) \), where $\bar{Y}_1$ is the mean of $Y$-values associated with the values of $M$ less than its median, and $\bar{Y}_2$ is that for the larger than median values of $M$. Bartlett followed the same selection criterion of $z_j$'s but suggested exclusion of middle 1/3 of the values, and his estimator is based on the lower and upper 1/3 of the values of $M$ and associated $Y$'s. The estimator is expressed as \( \hat{\beta}_{1B} = (\bar{Y}_3 - \bar{Y}_1)/(\bar{M}_3 - \bar{M}_1) \), where $Y_1$ is the mean of $Y$-values associated with the smallest 1/3 of the values of $M$, and $\bar{Y}_3$ is that for the largest 1/3. Durbin proposed to use the rank of $M$ as $z_j$'s. His method yields the following estimator of the slope parameter \( \hat{\beta}_{1D} = \left[ \sum_{j=1}^{n} j Y_j \right] \bigg/ \left[ \sum_{j=1}^{n} j m_j \right] \), but does not define the estimator of the intercept.

The IV method of estimation of the regression parameters does not require any unrealistic assumption on the reliability ratio. But the actual estimator depends on how the IV is defined, as definition of $Z$ affects both the estimator and its variance. This paper proposes a new method of defining IV based on the reflection of the explanatory or manifest variable. In general, the available methods of defining IV causes a significant loss of sample information (data) either by replacing the observed values of the explanatory variable by -1 or +1, or exclusion of some data, or due to ranking of data. But the proposed definition of the IV does not loss any information. Furthermore, the method produces more precise estimator than those proposed by Wald, Bartlett, and Durbin. Moreover, the new estimator based on the reflection of manifest variable is unbiased and consistent.

4 Proposed new IV and estimator

To avoid the unwanted and troublesome influence of the measurement error in the explanatory variable, the idea of reflection of the manifest variable is used for all the values of explanatory variable. The reflection of the points is taken about the fitted regression line. This is essentially done by a transformation of the observed values of the explanatory variable to their reflection on the Euclidean plane. In the conventional notation, the reflection of the explanatory variable $M_j = X_j + u_j$ (with measurement error $u_j$) for $j = 1, 2, \ldots, n$, can be defined as

\[
X^* = M \cos 2\psi + (Y - \hat{\beta}_{0m}) \sin 2\psi,
\]

where $\hat{\beta}_{0m}$ is the least square estimate of the intercept parameter, $\psi$ is the angle measure defined as $\psi = \arctan \hat{\beta}_{1m}$ in which $\hat{\beta}_{1m}$ is the least square estimate of the slope parameter in the manifest model, and $\cos$ and $\sin$ are the usual trigonometric cosine and sine functions respectively. For the definition of reflection points on the Cartesian plane readers may see Vaisman (1997, p. 164-169).

The proposed method requires to compute the reflection of all the data points, and use the transformed values of $M$, say, $X^*$ in defining the IV to fit the regression line of $Y$. The
estimator of the slope parameter under the proposed method is

$$\tilde{\beta}_{1R} = (Z_r'M)^{-1} Z_r'y = \frac{S_{x'y}}{S_{mm}}$$  \hfill (4.2)

where $Z_r = \begin{pmatrix} x_1^* \\ x_2^* \\ \vdots \\ x_n^* \end{pmatrix}$ and $S_{x'*m} = S_{mm}$ in which $S_{x'*m} = \sum_{j=1}^{n}(x_j^* - \bar{x})(m_j - m)$. It can be shown that Cov($Z_r, u) = Cov(Z_r, v) = 0$, that is, the proposed IV is independent of $u$ and $v$, but very strongly correlated with $M$. Also it can be easily shown that $E[M_j] = E[X_j] = E[X_j^*] = \mu_x$.

**Theorem 4.1** The estimator of the slope parameter of the simple regression model using IV based on the reflection of $M$ is the same as that produced by $X$, that is, $\tilde{\beta}_{1X} = \tilde{\beta}_{1R}$.

**Proof:** From the definition we get

$$\tilde{\beta}_{1X} = \frac{S_{xy}}{S_{xx}} = \frac{S_{xy}}{S_{xx}} \quad \text{and} \quad \tilde{\beta}_{1R} = \frac{S_{x'y}}{S_{mm}}$$  \hfill (4.3)

From (2.2), it is easy to show that $S_{my} = S_{xy}$ and $S_{mm} = S_{xx} + S_{uu}$. But the main body of the proof is based on

$$S_{x'y} - S_{my} = SSE_m \sin 2\psi,$$  \hfill (4.4)

where $\psi$ is as defined in equation (4.1), and $SSE_m$ is the sum of squares error for the manifest model. The above result follows from the fact that

$$x_j^* - m_j = m_j \cos 2\psi + (y_j - \tilde{y}_j) \sin 2\psi - m_j$$
$$= m_j (\cos 2\psi - 1) + y_j \sin 2\psi - \tilde{y}_j \sin 2\psi - \bar{m} \sin 2\psi$$
$$= -m_j (2 \sin^2 \psi) + y_j \sin 2\psi - \bar{y} \sin 2\psi + \bar{m} \sin 2\psi$$
$$= (y_j - \bar{y}) \sin 2\psi - (m_j - m) \sin 2\psi,$$  \hfill (4.5)

where $x_j^*$ is the reflection of $m_j$. Multiplying both sides of the above equation by $y_j$ and taking sum over $j$, we get

$$\sum (x_j^* - m_j)y_j = \sum (y_j - \bar{y}) y_j \sin 2\psi - \sum (m_j - \bar{m}) y_j \sin 2\psi$$
$$S_{x'y} - S_{my} = S_{yy} \sin 2\psi - S_{my} \sin 2\psi$$
$$\frac{S_{x'y} - S_{my}}{\sin 2\psi} = SST - SSR_m = SSE_m,$$  \hfill (4.6)

where $S_{yy} = SST$ is the sum of squares total, $SSR_m$ is the sum of squares regression, and $SSE_m$ is the sum of squares error for the regression of $Y$ on $M$. Note that $\frac{2 \sin^2 \psi}{\sin 2\psi} = \tan \psi = \tilde{\beta}_{1m}$. 
Instrumental variable estimator of the slope parameter…

Figure 1: Graph representing the Sum of Squares and Products in the presence of measurement error in the explanatory variable.

Then using equation (4.6), we can write

\[
\hat{\beta}_{1x} = \frac{S_{xy}}{S_{xx}} = \frac{S_{my} = S_{x'y} - SSE_m \sin 2\psi}{S_{mm} - S_{uu}}
\]

\[
\hat{\beta}_{1R} = \frac{S_{x'y}}{S_{mm}} = \frac{S_{my} + SSE_m \sin 2\psi}{S_{xx} + S_{uu}} = \frac{S_{xy} + SSE_m \sin 2\psi}{S_{xx} + S_{uu}}
\]

From Figure 2 \( \angle FAD = \angle FBE \) then

\[
\hat{\beta}_{1R} = \frac{S_{x'y}}{S_{mm}} = \frac{S_{x'y} - S_{xy}}{S_{mm} - S_{xx}} \tag{4.7}
\]

which leads to

\[
S_{x'y}S_{xx} = S_{xy}S_{mm} \tag{4.8}
\]

and finally simplification yields

\[
\frac{S_{x'y}}{S_{mm}} = \frac{S_{xy}}{S_{xx}} \text{ or, } \hat{\beta}_{1R} = \hat{\beta}_{1X} \tag{4.9}
\]

Hence the proof

4.1 Geometric Explanation

The presence of measurement error in the explanatory variable and its impact on the estimator of the slope as well as how the proposed method ‘treats’ the measurement error can be explained by graphs. The graphical representation also explains how the actual estimator of the slope is recovered by the new method.

Figure 1 represents the sum of squares and sum of products associated with the definition of the estimators of slope both for the latent and manifest variables. This graph represents
the presence of measurement error in the explanatory variable as well as the two estimators of the slope parameter. On the other hand Figure 2 displays the same along with that of the reflection of the manifest variable and three estimators of the slope parameter.

From Figure 1, the true estimator of the slope when the latent variable is available, that is, $\hat{\beta}_{1X}$ is represented by the tan of $\angle BAC$ of $\triangle ABC$. In the absence of the values of the latent variable this is unavailable. But for the manifest variable one can find the estimator of the slope to be $\hat{\beta}_{1m}$ which is represented by the tan of $\angle DAE$ of $\triangle ADE$. Note that here $DC$ (or equivalently $BE$) represents the sum of squares of measurement error ($S_{eu}$). Furthermore, under the assumptions of $E[Y|u] = 0$ and $E[X|u] = 0$, we have $BC = DE$ or $S_{xy} = S_{my}$. Finally, $\hat{\beta}_{1X} = \frac{S_{xy}}{S_{xx}} = \frac{BC}{AC}$, and $\hat{\beta}_{1m} = \frac{S_{my}}{S_{mm}} = \frac{ED}{AD}$.

The introduction of the reflection of the manifest variable changes $\triangle ADE$ of Figure 1 to $\triangle ADF$ in Figure 2. In fact the main difference between the two Figures is that Figure 2 has the small $\triangle BEF$ added to Figure 1. This triangle represents the effect of the reflection of the manifest variable. From Figure 2 the estimates of the slope are

$$\hat{\beta}_{1m} = \frac{S_{my}}{S_{mm}} \left(= \frac{DE}{DA}\right)$$

$$\hat{\beta}_{1X} = \frac{S_{xy}}{S_{xx}} \left(= \frac{BC}{AC}\right)$$

$$\hat{\beta}_{1R} = \frac{S_{xz,y}}{S_{mm}} \left(= \frac{FD}{AD}\right).$$

Since the tan of $\angle BAC$ represents the estimator $\hat{\beta}_{1X}$ and tan of $\angle DAF$ represents $\hat{\beta}_{1R}$, we conclude that $\hat{\beta}_{1X} = \hat{\beta}_{1R}$. Note that $\angle BAC = \angle DAF$.
5 Some properties and relationships

The estimated regression lines based on the OLS, and IV methods are summarised in the following way:

\[
\hat{Y}_X = \hat{\beta}_0X + \hat{\beta}_1X X \quad (5.1)
\]
\[
\hat{Y}_R = \hat{\beta}_0R + \hat{\beta}_1RM \quad (5.2)
\]
\[
\hat{Y}_W = \hat{\beta}_0W + \hat{\beta}_1WM \quad (5.3)
\]
\[
\hat{Y}_B = \hat{\beta}_0B + \hat{\beta}_1BM. \quad (5.4)
\]

Obviously, in the absence of \(X\), the fitted model in (5.1) is unavailable. The other fitted lines are obtainable since the manifest variable \(M\) is always observed along with the response \(Y\). Furthermore, even though the regression parameters are the same, the estimated models are different since observed \(M\) is different from the true value of the explanatory variable \(X\). Thus

\[
\hat{\beta}_0x + \hat{\beta}_1x X \neq \hat{\beta}_0z + \hat{\beta}_1z M.
\]

Another useful fact is that the sum of squares total is the same for regression of \(Y\) on \(X\) and that on \(M\). That is,

\[
SS_{yy} = SSR_x + SSE_x = SSR_m + SSE_m.
\]

Similarly, the following relationship of the regression sum of squares for models using \(X\), \(M\), and \(X^*\) are observed:

\[
SSR_x = \hat{\beta}_1xSS_yx = \hat{\beta}_1RSS_{my} = SSR_R \neq \hat{\beta}_1RSS_{mm} = \hat{\beta}_1RS_{x*y}.
\]

Finally, the coefficient of determination is noted to be

\[
R^2_x = \frac{SSR_x}{SS_T^2} = \frac{SSR_R}{SS_T^2}.
\]

6 Examples for Illustration

In this section, two illustrative examples based on two real life data sets are provided. Both cases reveal the superiority of the proposed new IV estimator. The first data set has measurement error in the explanatory variable only, but the second data set has measurement error in both the response. For the second example assume that the ratio of error variance \(\lambda = \frac{\sigma_{ee}}{\sigma_{uu}} < 1\), where \(\sigma_{ee}\) is the error variance of the response variable and \(\sigma_{uu}\) is the error variance of the explanatory variable.

6.1 Yield of Corn Data

The data set of the first example deals with the yield of corn \((Y)\) for different levels of soil nitrogen \((M)\), and is taken from Fuller (2006, p.18). Here the explanatory variabl
soil nitrogen level, has been determined with measurement error. Fuller has analysed the data with the existing method with usual assumptions including known reliability ratio. We provide the regression analyses of the data for both with (a) the measurement error in the explanatory variable $M$, and (b) the instrumental variables including one defined by $X^*$, the reflection of the observed explanatory variable $M$. Comparison of the regression estimates and related results from different methods are provided below. The Table 1 below shows the fitted regression lines, mean sum of squares error, and the coefficient of determination based on the OLS and various IV methods including the reflection method.

<table>
<thead>
<tr>
<th>Method</th>
<th>Fitted regression equation</th>
<th>MS Error</th>
<th>$R^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Least Squares</td>
<td>$\hat{Y}_M = 73.153 + 0.344M$</td>
<td>57.321</td>
<td>0.412</td>
</tr>
<tr>
<td>Wald</td>
<td>$\hat{Y}_W = 75.91 + 0.305M$</td>
<td>60.98</td>
<td>0.364</td>
</tr>
<tr>
<td>Bartlett</td>
<td>$\hat{Y}_B = 72.38 + 0.355M$</td>
<td>56.05</td>
<td>0.425</td>
</tr>
<tr>
<td>Reflection</td>
<td>$\hat{Y}_R = 65.8164 + 0.4479M$</td>
<td>45.224</td>
<td>0.536</td>
</tr>
<tr>
<td>$\sigma_{uu}$ Known</td>
<td>$\hat{Y}_V = 67.561 + 0.423M$</td>
<td>48.125</td>
<td>0.506</td>
</tr>
</tbody>
</table>

The straightforward regression of $Y$ on $M$ produces the estimated (OLS) regression line, $\hat{Y}_m = 73.153 + 0.344M$ with mean sum of squares error, $MSE_m = 57.321$ (see the first regression line in Table 1) and $R^2_m = 0.421$. This analysis does not take into account the presence, and hence the effect, of the measurement errors in the explanatory variable. As such these results are not based on any sound statistical method and hence unacceptable.

Fuller (2006, p.18-19) assumes that $\sigma_{uu} = 57$, and that the reliability ratio, $\kappa_{ym}$, is known. Under the above assumptions the estimated regression line reported to be $\hat{Y}_V = 67.561 + 0.423M$ with modified $MSE$, $MSE_V = \hat{\sigma}_{se} = 48.125$, and $R^2_V = 0.506$. Clearly, there has been an improvement in the proportion of variability in $Y$ that is explained by $M$ under the method used by Fuller (2006). The MSE has also decreased (from $MSE_m = 57.321$ to $MSE_V = 48.125$) under the Fuller method. Thus the Fuller method is not only a better method than the OLS, but also provides a much better fit.

The use of the reflection of $M$ in the specification of the instrumental variable leads to the fitted regression line, $\hat{Y}_R = 65.8164 + 0.4479M$ with mean sum of squares error, $MSE_R = 45.224$ (see second last row of Table 1) and $R^2_R = 0.536$. Unlike Fuller's method, these results are obtained without additional assumptions on any of the parameters of the model or the reliability ratio. However, the regression parameters obtained by using the reflection of $M$ are fairly close to those obtained by Fuller under the previously stated assumptions. The regression line produced by the proposed method provides a much better fit than that obtained by Fuller. Obviously, the $MSE_R$ under the IV is much smaller than $\hat{\sigma}_{se}$ obtained by Fuller's method. Moreover, under the proposed method the value of the coefficient of determination is 53.6%, compared to only 50.60% under the Fuller's method.

The estimates of the regression parameters of the manifest model are $\hat{\beta}_{1m} = 0.344$ and $\hat{\beta}_{0m} = 73.153$, and that of the proposed instrumental variable model are $\hat{\beta}_{1R} = 0.4479$ and $\hat{\beta}_{0R} = 65.8164$. These figures support the results in Theorem (4.1), that is, $\hat{\beta}_{1R} = 0.4479 >$
\[ \hat{\beta}_{1m} = 0.344, \text{ and } \hat{\beta}_{0R} = 65.8164 < \hat{\beta}_{0m} = 73.153. \text{ Note here the correlation is positive.} \]

It is important to compare the results of the new IV estimator with other IV estimators such as the Wald and Bartlett methods specified earlier. The results of Wald method yields, \( \hat{Y}_W = 75.91 + 0.305M \) with \( MSE_W = 60.98 \) and \( R^2_W = 0.364 \). Moreover, using Bartlett’s definition of the IV, we get \( \hat{Y}_B = 72.38 + 0.355M \) with \( MSE_B = 56.05 \) and \( R^2_B = 0.425 \). Practically both methods are inefficient, although the Bartlett method produces better fit (larger \( R^2_B \)) than that of Wald \( R^2_W \).

Clearly, the Wald’s method produces the worst of the five fitted models in terms of the MSE (or \( R^2 \)). The Bartlett’s method is better than the Wald’s method in terms of the value of \( R^2 \). The Fuller’s method provides a much better fit than the OLS, Wald and Bartlett methods. However, the reflection based IV fitted model has the largest \( R^2 \). At the same time the regression estimates of the slope and intercept for the Fuller method is much close to that of the reflection based estimator. Thus the IV based on the reflection of \( M \) provides the best model without making any additional assumptions on the error variance or reliability ratio.

### 6.2 Hen Pheasants Data

The data set for the second example is also taken from Fuller (2006, p.34). The data deal with the number of hen pheasants in Iowa at two different season/time of the year, and were collected by the Iowa Conservation Commission. These data are based on the average number of birds sighted by trained observers traveling a number of specific routes in late April and early May, and again in August. Both measures are subject to error for two reasons. First, the routes are a sample of all possible routes in Iowa. Second, observers cannot be expected to sight all pheasants along the route. The response variable \( Y \) is the average number of hens in August, and the explanatory variable \( M \) is the average number of hens in Spring, where the ratio of error variances \( \lambda < 1 \). On the basis of previous analyses, it has been estimated that the error variance for the Spring count is about six times larger than that in August. The fitted regression models and associated statistics are provided in the Table 2 below.

<table>
<thead>
<tr>
<th>Method</th>
<th>Fitted regression equation</th>
<th>MS Error</th>
<th>( R^2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Least Squares</td>
<td>( \hat{Y}_M = 2.142 + 0.649M )</td>
<td>0.347</td>
<td>0.826</td>
</tr>
<tr>
<td>Wald</td>
<td>( \hat{Y}_W = 2.498 + 0.614M )</td>
<td>0.44</td>
<td>0.78</td>
</tr>
<tr>
<td>Bartlett</td>
<td>( \hat{Y}_B = 2.036 + 0.66M )</td>
<td>0.32</td>
<td>0.84</td>
</tr>
<tr>
<td>Reflection</td>
<td>( \hat{Y}_R = 1.323 + 0.731M )</td>
<td>0.14</td>
<td>0.93</td>
</tr>
<tr>
<td>Moments</td>
<td>( \hat{Y}_{MO} = 1.116 + 0.751M )</td>
<td>0.09</td>
<td>0.95</td>
</tr>
</tbody>
</table>

The first regression equation and the associated statistics in Table 2, \( \hat{Y}_m = 2.142 + 0.649M, MSE_m = 0.347 \) and \( R^2_m = 0.826 \), are obtained by the OLS method using \( M \) which is subject to the measurement error.
The method of moments (MOM) estimator, under the assumption that the ratio \( \delta = \sigma_{oe}^{-1} \sigma_{ee} \) is known can be found. Following Fuller (2006, p.35), for \( \delta = \frac{1}{8} \), the fitted regression equation becomes \( \hat{Y}_{MO} = 1.1158 + 0.7516M \) with \( MSE_{MO} = 0.09 \) and \( R^2_{MO} = 0.95 \). This is a much better fitted model, with an increased value of \( R^2 \), than that obtained by the OLS method.

The second last row of Table 2 represents the regression line and other statistics produced by the proposed new instrumental variable method based on the reflection of \( M \): \( \hat{Y}_R = 1.323 + 0.731M \), \( MSE_R = 0.130 \) and \( R^2_R = 0.93 \).

The IV estimator based on Wald’s method yields \( \hat{Y}_W = 2.498 + 0.614M \) with \( MSE_W = 0.44 \) and \( R^2_W = 0.78 \). Similarly, Bartlett’s IV method gives \( \hat{Y}_B = 2.036 + 0.66M \), \( MSE_B = 0.32 \) and \( R^2_B = 0.84 \).

In terms of the \( R^2 \) value the Wald’s method is the worst, followed by the OLS method. Thus Wald’s IV method may produce worst fit than the OLS method. The Bartlett’s method gives a similar \( R^2 \) as the OLS method. However, the MOM estimation produces the largest \( R^2 \), although it is not too far from that produced by the proposed reflection based method. It is important to note that the MOM is based on the assumption that the value of \( \delta \) is known, whereas no assumption is required for the reflection method. Furthermore, due to the nature of the definition of the IV, we have only ‘treated’ the measurement error in the explanatory variable. It seems that similar treatment of the response variable would produce results better than the method of moments.

Among the IV estimators the proposed reflection based IV performs much better than the others in terms of providing the best fitted model with largest \( R^2 \). This is not surprising due to the fact that IVs proposed by Wald, Bartlett, and Durbin fails to use part of the information of the sample data to define the IV. Although the MOM estimation method provides slightly better fit than the proposed reflection based IV method, the former is dependent on the unrealistic assumption that \( \delta \) is known. When no information is available on \( \delta \), which is normally the case, the new method ensures the best fitted model.

## 7 Concluding Remarks

The paper considers the simple regression model with measurement error in the explanatory variable. It proposes a new estimation procedure based on the idea of a new instrumental variable which is defined from reflection of the manifest variable. Also, it provides the theory of the available literature and compares the existing methods with proposed new method. Unlike, some of the existing methods it does not loss of information. Moreover, the statistical properties of the proposed estimator are much superior to those of the available methods in the literature.

The analytical results and the illustrative examples demonstrate the fact that the proposed method significantly reduces the mean sum of squares error than the currently used methods. As such, the coefficient of determination of the proposed method is higher than
that of the existing methods.

The proposed method in the paper is new, easy to implement, and performs much better than the existing method. We have demonstrated the superiority of our method both analytically and via numerical illustration.

Surprisingly, the proposed IV method recovers the true estimator of the slope, $\hat{\beta}_{1X}$, from the manifest variable and stochastic model even if the true values of the latent variable is unobservable. The Theorem 4.1 and the Figure 2 demonstrate this remarkable fact. The same comment would apply for the estimator of the intercept.

References


Bayesian Prediction of Matrix Elliptical Multivariate Models with Conjugate Prior

Mohammad Arashi $^1$ and Shahjahan Khan $^2$

$^1$Faculty of Mathematics, Shahrood University of Technology
P.O. Box 316-3619995161, Shahrood, Iran
Email: m_arashi.stat@yahoo.com

$^2$Department of Mathematics and Computing
Australian Centre for Sustainable Catchments
University of Southern Queensland
Toowoomba, Queensland, AUSTRALIA,
Email: khans@usq.edu.au

Abstract

This paper considers a multivariate regression model with a matrix variate elliptically contoured (MEC) distribution for the responses. The MEC is defined as a mixture distribution of inverse Laplace transformation and matrix variate normal distribution. The prediction distribution of a set of matrix future responses from the same model with common indexing parameters is obtained under Bayesian framework with the conjugate prior of the parameters. The prediction distribution is found to be a matrix-T distribution. The results of the paper is a generalization of earlier results for linear models in terms of the generalization of the (i) multiple regression model, (ii) normal or Student-t distribution, and (iii) joint conjugate prior distribution.

Key words and phrases: Multivariate regression model, Elliptically contoured distribution, Inverse Laplace transform, Conjugate prior, matrix-T distribution, and Prediction.

1 Introduction and Some Preliminaries

Prediction analysis (see Aitchison and Dunsmore 1975, and Geisser 1993) is predominantly a Bayesian method. The predictive inference is usually based on the prediction distribution of
Bayesian Prediction of Matrix Elliptical Multivariate Models

future unobserved responses, conditional on the realized responses from the same model. The derivation of prediction distribution involves integration of the parameters from the product of the joint distribution of the realized and future responses and that of the underlying parameters. In the multivariate case such integrations are over the multidimensional spaces of Manifolds.

Under the Bayesian framework, there has been many systematic studies for linear regression models with non-normal errors. Interested readers may refer to the papers by Zellner (1976), Singh et al. (1995), Jammalamadaka et al. (1987), Chib et al. (1991), Khan and Haq (1994), Osiewalski and Steel (1993), and more recently Fang and Li (1999), Khan (2002, 2004), Ng (2002), Arashi (2010), Ng (2010), Vidal and Arellano-Valle (2010), and Tsukuma (2010).

In this section, we propose some necessary tools for mathematical computations with matrices. Also this paper we consider a wider class of distributions namely the matrix variate elliptically contoured (MEC) distributions. It is well known that a large number of commonly used distributions are member of the MEC family including the most popular matrix normal and matrix-T distributions.

Let $X$ be an $n \times p$ random matrix, which can be expressed in terms of its elements, columns and rows as

$$X = (x_{ij}) = (x_1, \ldots, x_p) = (x_{1(1)}, \ldots, x_{n(1)})'. \quad (1.1)$$

Here $x_{1(1)}, \ldots, x_{n(1)}$ can be regarded as a sample of size $n$ from a $p$-dimensional population.

As pointed by Fang and Li (1999), there are many ways (not completely different) to define matrix variate elliptically contoured distributions. Four classes of elliptical matrix variate distributions are defined and discussed by Dawid (1977), and Anderson and Fang (1990). For the purpose of this paper we specifically consider the following situation

**Definition 1.1.** An $n \times p$ random matrix $X$ has a family of MEC distributions if its density has the form

$$g(X) = d_{n,p}|\Sigma|^{-\frac{p}{2}} f \{\Sigma^{-1}(X - 1\mu')(X - 1\mu')\}, \quad (1.2)$$

where $\mu \in \mathbb{R}^p$ and $\Sigma$ is a $p \times p$ semi-positive definite matrix. This distribution is denoted by $X \sim E_{n,p}(\mu, I_n \otimes \Sigma, f)$. For notational convenience we may also use $X \sim E_{n,p}(\mu, I_n, \Sigma, f)$ where needed.

Definition 1.1 imposes the condition $f(AB) = f(BA)$ on the density generator $f$ for any $p \times p$ positive definite symmetric matrices $A$ and $B$. Throughout the paper, without
loss of generality, we take $tr$ operation from the argument of $f(.)$.
Some commonly used members of the MEC distributions are stated below.

(i) Matrix Variate Normal (MN) Distribution
A random matrix $X \in \mathbb{R}^{n \times p}$ has a MN distribution, with mean $M$, row and column covariance matrices $\Omega$ and $\Sigma$ respectively is denoted by $X \sim MN_{n,p}(M, \Sigma, \Omega)$, if its pdf is given by

$$f(X) = \frac{|\Omega|^{-\frac{\nu}{2}}|\Sigma|^{-\frac{\nu}{2}}}{(2\pi)^{\frac{np}{2}}} \exp\left\{ -\frac{1}{2} \text{tr} \left[ \Omega^{-1}(X - M)\Sigma^{-1}(X - M) \right] \right\}.$$

(ii) Matrix Variate Student-t (MT) Distribution
A random matrix $X \in \mathbb{R}^{n \times p}$ has a MT distribution, with mean $M$, row and column scale matrices $\Omega$ and $\Sigma$ respectively and $\nu$ d.f. is denoted by $X \sim MT_{n,p}(M, \Sigma, \Omega, \nu)$, if its pdf is given by

$$f(X) = \frac{|\Omega|^{-\frac{\nu}{2}}|\Sigma|^{-\frac{\nu}{2}}}{g_{n,p}} \left| I_n + \Omega^{-1}(X - M)\Sigma^{-1}(X - M) \right|^{-\frac{n+p+\nu-1}{2}},$$
where

$$g_{n,p} = \frac{(\nu\pi)^{\frac{\nu}{2}}\Gamma_p\left(\frac{\nu+p-1}{2}\right)}{\Gamma_p\left(\frac{\nu+n+p-1}{2}\right)}.$$

(iii) Matrix Variate Pearson Type-VII (MPVII) Distribution
A random matrix $X \in \mathbb{R}^{n \times p}$ has a MPVII distribution, with mean $M$, row and column scale matrices $\Omega$ and $\Sigma$ respectively and parameters $m$ and $q$ is denoted by $X \sim MPVII_{n,p}(M, \Sigma, \Omega, m, q)$, if its pdf is given by

$$f(X) = \frac{|\Omega|^{-\frac{m}{2}}|\Sigma|^{-\frac{q}{2}}}{h_{n,p,m,q}} \left| I_n + \Omega^{-1}(X - M)\Sigma^{-1}(X - M) \right|^{-m},$$
where

$$h_{n,p,m,q} = \frac{(q\pi)^{\frac{q}{2}}\Gamma_p\left(m - \frac{q}{2}\right)}{\Gamma_p\left(m\right)}.$$

(iv) Matrix Variate Power Exponential (MPE) Distribution
A random matrix $X \in \mathbb{R}^{n \times p}$ has a MPE distribution, with mean $M$, row and column scale matrices $\Omega$ and $\Sigma$ respectively and parameters $r$ and $s$ is denoted by $X \sim MPE_{n,p}(M, \Sigma, \Omega, r, s)$, if its pdf is given by

$$f(X) = \frac{|\Omega|^{-\frac{r}{2}}|\Sigma|^{-\frac{s}{2}}}{f_{n,p,r,s}} \exp\left\{ -\frac{r}{2} \left( \text{tr} \left[ \Omega^{-1}(X - M)\Sigma^{-1}(X - M) \right] \right) ^s \right\},$$
where
\[ j_{n,p,r,s} = \frac{(2\pi)^{p/2} \Gamma \left( \frac{n+p}{2} \right)}{s^{n/2} \Gamma \left( \frac{n}{2} \right) r^{np/2}}. \]

(v) Matrix Variate Laplace (ML) Distribution

A random matrix \( X \in \mathbb{R}^{n \times p} \) has a ML distribution, with mean \( M \), row and column scale matrices \( \Omega \) and \( \Sigma \) respectively is denoted by \( X \sim ML_{n,p}(M, \Sigma, \Omega) \), if its pdf is given by
\[ f(X) = \frac{|\Omega|^{-\frac{n}{2}} |\Sigma|^{-\frac{p}{2}}}{I_{n,p}} \exp \left\{ -\frac{r}{2} \left( \text{tr} \left[ \Omega^{-1}(X - M)' \Sigma^{-1}(X - M) \right] \right) \right\}, \]
where
\[ I_{n,p} = \frac{2\pi^{np/2} \Gamma \left( \frac{np}{2} \right)}{\Gamma \left( \frac{n}{2} \right) \Gamma \left( \frac{p}{2} \right)}. \]

(vi) Matrix Variate Kotz-Type (MK) Distribution

A random matrix \( X \in \mathbb{R}^{n \times p} \) has a MK distribution, with mean \( M \), row and column scale matrices \( \Omega \) and \( \Sigma \) respectively is denoted by \( X \sim MK_{n,p}(M, \Sigma, \Omega, \delta) \), if its pdf is given by
\[ f(X) = \frac{|\Omega|^{-\frac{n}{2}} |\Sigma|^{-\frac{p}{2}}}{I_{n,p,r}} \exp \left\{ -\frac{r}{2} \left( \text{tr} \left[ \Omega^{-1}(X - M)' \Sigma^{-1}(X - M) \right] \right) \right\}, \]
where
\[ I_{n,p,r} = \frac{(2\pi)^{np/2} \Gamma \left( \frac{np}{2} \right)}{\Gamma \left( \frac{n}{2} \right) \Gamma \left( \frac{p}{2} \right) \Gamma \left( \frac{np}{2} \right) r^{np/2}}. \]

It is well known that the family of MEC distributions can always be expressed in an integral form of a matrix variate normal distribution with weight function as given in the following theorem.

Theorem 1.1. Let \( X \sim E_{n,p}(\mu, I_n \otimes \Sigma, f) \) where the pdf \( g(X) \) of \( X \) is defined by
\[ g(X) = |\Sigma|^{-n/2} h \left[ \text{tr} \Sigma^{-1}(X - 1\mu)'(X - 1\mu) \right]. \]
If \( h(t) \), \( t \in [0, \infty) \) has the inverse Laplace transform (denoted by \( \mathcal{L}^{-1}[h(t)] \)), then we have
\[ g(X) = \int_0^{\infty} w(z) f_N(\mu, z^{-1}, I_n \otimes \Sigma)(X) \, dz, \quad (1.3) \]
where $f_{N_{n,p}(\mu, z^{-1} I_n \otimes \Sigma)}(X)$ stands for the pdf of the $n \times p$ matrix $X$ distributed as matrix variate normal with the mean matrix $\mu'$ and the covariance matrix $z^{-1} I_n \otimes \Sigma$, and $w(z)$ is the weight function given by

$$w(z) = (2\pi)^{np/2} z^{-np/2} L^{-1}[h(2z)].$$

(1.4)

For the proof refer to Theorem 4.2.1 of Gupta and Varga (1995).

**Remark 1.1.** In Theorem 1.1 it is stated that $f$ is the density of $N_{n,p}(\mu, z^{-1} I_n \otimes \Sigma)$. It is realized from the proof that $f$ can even has one of the following densities

1. $N_{n,p}(\mu, z^{-1} I_n, \Sigma)$ or

2. $N_{n,p}(\mu, I_n, z^{-1} \Sigma)$ or

3. $N_{n,p}(\mu, z^{-1/2} I_n, z^{-1/2} \Sigma)$.

The above facts enable us to adopt each representation whenever is needed for practical use. In sequel we present some necessary tools for the derivations in forthcoming sections.

Section 2 describes the multivariate regression model as a generalization of the multiple regression model. The associated future multivariate regression model as well as the prediction distribution are covered in Section 3. Some concluding remarks are provided in Section 4.

## 2 Multivariate Regression Model

In this section, we consider a multivariate regression model with the MEC errors. For a precise setup consider the following regression model

$$Y = BX + E,$$

(2.1)

where the $n$ columns of the $p \times n$ response matrix $Y$ can be regarded as a sample of size $n$ from a $p$-dimensional population. $X$ is the $k \times n$ design matrix of known values of rank $k$, $B$ is the $p \times k$ matrix of unknown regression parameters, $n > p + k$. The $p \times n$ error matrix $E$ is assumed to have a MEC distribution, $E_{p,n}(0, \Phi^{-1} \otimes I_n, f)$.

From Theorem 1.1 we can immediately write

$$f(Y|B, \Phi) = \int_0^{\infty} f_{N(BX, t^{-1} \Phi^{-1} \otimes I_n)}(Y) W^*(t) dt$$

$$= \int_0^{\infty} f(Y|B, \Phi, t) W^*(t) dt,$$

(2.2)
where $W^*(t)$ is the weight matrix. Now following Ng (2010) let $\Sigma = t\Phi$ with the Jacobian of transformation $J(\Sigma \rightarrow \Phi) = t^{p+k}$ \frac{1}{2}. Further adopt a normal-Wishart conjugate prior for $(B, \Sigma)$ as

$$
\pi(B|\Sigma) \propto |\Sigma|^\frac{3}{2} \text{etr}\left\{-\frac{1}{2} \Sigma(B - B^*)Y(B - B^*)'\right\},
$$

$$
\pi(\Sigma) \propto |\Sigma|^{\nu-k-1} \text{etr}\left\{-\frac{1}{2} \Omega \Sigma\right\},
$$

where $\nu, p \times k$ matrix $B^*$, $k \times k$ matrix $Y$ and $p \times p$ matrix $\Omega$ are all known hyperparameters.

Suppose

$$
\pi(B, \Phi|t) \propto \pi(B|\Sigma)\pi(\Sigma)J(\Sigma \rightarrow \Phi)
\propto t^{\frac{(p+k)}{2}} |\Phi|^{\nu-k-1} \text{etr}\left\{-\frac{1}{2} t\Phi [(B - B^*)'Y(B - B^*) + \Omega]\right\}. \quad (2.3)
$$

Then the conjugate prior distribution for the MEC distribution can be obtained as

$$
\pi(B, \Phi) \propto \int_0^\infty \pi(B, \Phi|t)W^*(t)dt. \quad (2.4)
$$

### 3 Prediction Distribution

Suppose $Y$ in equation (2.1) is observable and $Y_f$ in

$$
Y_f = BX_f + E_f \quad (3.1)
$$

is an observable $p \times n_f$ matrix of future responses with a $k \times n_f$ design matrix $X_f$ of known values with rank $k$, satisfying $n + n_f > p + k$.

In the same fashion as in Khan (2002) and Ng (2010), the joint density of $(Y, Y_f)$ is then given by

$$
f(Y, Y_f|B, \Phi) \propto \int_0^\infty f(Y, Y_f|B, \Phi, t)W^*(t)dt,
$$

where

$$
f(Y, Y_f|B, \Phi, t) \propto |t\Phi|^{\frac{1}{2}(n+n_f)} \text{etr}\left\{-\frac{1}{2} t\Phi \left\|Y - BX_f\right\|^2 + \left\|Y_f - BX_f\right\|^2\right\}. \quad (3.2)
$$

Consequently the Bayesian predictive density function of $Y_f$ is defined as

$$
f(Y_f|Y) \propto \int \int f(Y, Y_f|B, \Phi, t)\pi(B, \Phi|t)W^*(t)dBd\Phi dt, \quad (3.3)
$$

where $\pi(B, \Phi|t)$ is given by (2.3). In the sequel we propose the main result of this approach.
Theorem 3.1. For the multivariate regression model (2.1) where $Y$ follows a family of MEC distributions, using the conjugate prior on $(B, \Phi)$ given by (2.4), the predictive distribution of $Y_f$, is the matrix-$T$ distribution denoted by $Y_f|Y \sim T_{p,n_f}(M, H^{-1}, R, n_f + p + \nu - 1)$ with the following density

$$f(Y_f|Y) = c(p, n_f, \nu)|H|^{\frac{p}{2}}|R|^{-\frac{p}{2}}$$

$$\times |I_p + R^{-1}(Y_f - M)'H(Y_f - M)|^{-\frac{1}{2}(p+n_f+\nu-1)},$$

where

$$c(p, n_f, \nu)^{-1} = \left[ \Gamma \left( \frac{1}{2} \right) \right]^{p n_f} \frac{\Gamma_p \left[ \frac{1}{2} (n_f + \nu - 1) \right]}{\Gamma_p \left[ \frac{1}{2} (p + n_f + \nu - 1) \right]},$$

$$L = XX' + X_f X_f' + Y,$$

$$H = I_{n_f} - X_f L^{-1} X_f,$$

$$M = (YX' + B^*Y)L^{-1}X_f H^{-1},$$

$$R = YY' + B^*YY'B^* + \Omega + (YX' + B^*Y)L^{-1}(YX' + B^*Y)'$$

$$- (YX' + B^*Y)L^{-1}X_f M'$$

(3.4)

in which $\Gamma_p(\cdot)$ is the generalized gamma function defined as $\Gamma_p \left( \frac{1}{2} \right) = n^\frac{b(k-1)}{2} \prod_{i=1}^{b} \Gamma \left( \frac{a_i + 1}{2} \right)$.

Proof: Let $f(Y_f|t)$ be the density of $Y_f$ under normal assumption, then by definition we have

$$f(Y_f|Y) \propto \int f(Y_f|Y, t)W^*(t)dt,$$

where

$$f(Y_f|Y, t) = \int_{\Phi > 0} \int f(Y, Y_f|B, \Phi, t)\eta(B, \Phi|t)dBd\Phi$$

$$\propto \int \int \frac{1}{2}^\frac{p(n + n_f + \nu + k)}{2} |\Phi|^\frac{n_f + \nu + k}{2}$$

$$\times \text{etr}\left\{-\frac{1}{2} t\Phi \left[ \|Y - BX\|^2 + \|Y_f - BX_f\|^2 \right]\right\}$$

$$\times \text{etr}\left\{-\frac{1}{2} t\Phi \left[ (B - B^*)Y(B - B^*)' + \Omega \right]\right\} dBd\Phi.$$

(3.5)

Using the fact that

$$\|Y - BX\|^2 + \|Y_f - BX_f\|^2 + (B - B^*)Y(B - B^*)' + \Omega$$

$$= (Y_f - M)'H(Y_f - M)' + (B - \tilde{B})L(B - \tilde{B})' + R,$$
where
\[
\hat{B} = (Y_f'X_f' + B(XX' + Y))L^{-1}.
\]

Then the expression in (3.5) simplifies to
\[
f(Y_f|Y, t) \propto \int_{\Phi > 0} t^{\frac{1}{2}p(n+p_f+\nu)} |\Phi|^{\frac{n+n_f+\nu}{2} - \frac{p+1}{2}} \times \text{etr} \left\{ -\frac{t}{2} \Phi \left[ (Y_f - M)H(Y_f - M)' + R \right] \right\} d\Phi
\]
\[
\times t^{\frac{p+1}{2}} |\Phi|^{\frac{1}{2}} \text{etr} \left\{ -\frac{t}{2} \Phi (B - \hat{B})L(B - \hat{B})' \right\} dB.
\]

Integrating over \(B\), we get
\[
f(Y_f|Y, t) \propto \int_{\Phi > 0} t^{\frac{1}{2}p(n+n_f+\nu)} |\Phi|^{\frac{n+n_f+\nu}{2} - \frac{p+1}{2}} \times \text{etr} \left\{ -\frac{t}{2} \Phi \left[ (Y_f - M)H(Y_f - M)' + R \right] \right\} d\Phi
\]

Now make the transformation
\[
Z = t^{\frac{1}{2}} \Phi \left[ (Y_f - M)H(Y_f - M)' + R \right]
\]
with the Jacobian of the transformation \(|Z|^{-\frac{p+1}{2}}\) to get
\[
f(Y_f|Y, t) \propto \int_{Z > 0} t^{\frac{1}{2}p(n+n_f+\nu)} |(Y_f - M)H(Y_f - M)' + R|^{-\frac{p+1}{2}} t^{-1} Z^{\frac{n+n_f+\nu}{2} - \frac{p+1}{2}} \times |Z|^{-\frac{p+1}{2}} \text{etr} \left( -\frac{t}{2} Z \right) dZ
\]
\[
\propto |(Y_f - M)H(Y_f - M)' + R|^{-\frac{p+1}{2}} \times \int_{Z > 0} |Z|^{\frac{n+n_f+\nu}{2} - \frac{p+1}{2}} \text{etr} \left( -\frac{1}{2} Z \right) dZ
\]
\[
\propto |(Y_f - M)H(Y_f - M)' + R|^{-\frac{p+1}{2}}.
\]

Thus we conclude that
\[
f(Y_f|Y) \propto |(Y_f - M)H(Y_f - M)' + R|^{-\frac{p+1}{2}} \int W(t) dt
\]
\[
\propto |(Y_f - M)H(Y_f - M)' + R|^{-\frac{p+1}{2}}
\]
\[
\propto |I_p + R^{-1} [(Y_f - M)'H(Y_f - M)]|^{-\frac{p+1}{2}}.
\]
Remark 3.1. The predictive distribution of $Y_f$ under matrix normal responses is identical to the matrix-T distribution above (see Broemeling 1985, p.379). Thus one can carry out further inference for $Y_f$ and find the highest posterior density (HPD) region of $Y_f$ similar to the result of section 8.4 of Box and Tiao (1992).

4 Concluding Remarks

The paper considers a multivariate conjugate prior as a mixture of matrix normal and Wishart distributions for the location and scale parameter matrices. All previous work in the area within the Bayesian framework are based on non-informative prior involving only the scale parameters. Thus in addition to dealing with the generalized multivariate regression model the paper assumes a wider set of matrix prior distributions for both the location and scale parameters. As such the results of the papers can be applied for any simple and multiple regression model as special cases. Also, the distributional assumption of the paper, namely the MEC family of distributions, covers a large number of popularly used multivariate distributions. In effect the results of this paper are applicable to any member of the elliptically contoured family of distributions including the normal and Student-t distributions for both vector and matrix variate cases.

It is interesting to note that the prediction distribution of the matrix of future responses, conditional on the matrix of realized responses, is a matrix variate T distribution. From the properties of the matrix-T distribution (see Khan 2002) the prediction distribution of any sub-matrix of future responses, conditional on the relevant sub-matrix of realized responses, also follows a matrix-T distribution. Likewise, prediction distribution of any row (or column) vector of the future response matrix will be a multivariate Student-t distribution. It is important to note that the prediction distribution of any row (or column) of the future response matrix is not independent of other rows (or columns). So the rows (or columns) of the future response matrix are dependent, even though they may be uncorrelated.

Acknowledgements

First author's research was supported by Shahrood University of Technology.

REFERENCES


MODELING THE DAILY RAINFALL AMOUNT OF DIR AND D.I. KHAN, PAKISTAN FOR AGRICULTURAL PLANNING

Zahid Hussain
Department of Mathematics, Statistics & Computer Science
Khyber Pakhtunkhwa Agricultural University Peshawar
Email: zahidh54@yahoo.com

ABSTRACT

In the present study Gamma distribution models were used for modeling rainfall amount on seasonal basis to derive results of potential importance to agriculture. The daily rainfall data of two sites from Khyber Pakhtunkhwa province, Pakistan, viz. Dir and D.I. Khan (having quite significantly different mean annual rainfall & weather condition) were analyzed. The seasonal pattern of rainfall amount is summarized by fitting smoothed regression curves using generalized linear models. This enables to compute the mean amount of rain per rainy day for each day of the entire season/year, using the estimates of only a few parameters of the model. The adequacy of the fitted models for both the sites was tested by the analysis of deviance residuals & found satisfactory. The results reveal that Fourier series with 4 harmonics & 3 harmonics are suitable to fit the Gamma distribution for modeling mean rain per rainy day for Dir & D.I. Khan, respectively. Also pattern of rainfall amount (throughout the year) for both the sites is found to be bimodal.

This paper aims to show that with recent development in statistical methodology and the associated computing facilities have made it possible to use essentially the same range of methods for analyzing non-normally distributed data (such as daily rainfall data that is clearly non-normal) that was previously used only for normal data.

KEY WORDS

Agricultural planning; Gamma distribution; Generalized linear models; seasonal changes; Fourier series; deviance; deviance residuals.

1. INTRODUCTION

A comprehensive statistical analysis of daily rainfall data is of vital importance. Rainfall or the lack of it at particular times plays an important role in different spheres of activities such as meteorology, climatology, hydrology and agriculture. However, the main use of these analyses can be made in the field of agriculture since rainfall is the most important non-cash input responsible for maximum crop yields, particularly, in the rainfed and partially rainfed areas. In many parts of the world a detailed knowledge of rainfall regime is an important prerequisite for agricultural planning, e.g. for pre-sowing cultivation and for sowing. The information on various aspects of rainfall like the probability of rainfall and the distribution of rainfall amount throughout the year etc. is much vital to the farmers and agriculturists of the area concerned.
Daily rainfall data are clearly non-normally distributed. However, statistical methods particularly in the application of generalized linear models has much improved the range of techniques available for analysis of data that are non- normally distributed. In this study, the work was carried out using APL2 (A programming language 2). The availability of computer packages facilitates to fit and test a wide range of models for daily rainfall data.

Usually, probability models are taken into account for modeling the amount of rain on rainy days. In this regard, one common approach is to assume that the amount of rain on successive rainy days are independent and to fit some theoretical distribution to these amount under this assumption. It has been observed that the frequency distribution of rainfall amount has a J-shaped distribution and is highly skewed (positive) with a large frequency of small amount of rain and a small frequency of very large amount. In this context, an appropriate choice is the gamma distribution that usually illustrates the skewness of rainfall amount well.

In the present analysis, the gamma distribution with parameters varying with time has been fitted to the daily rainfall data (non-zero amount of rain) of Dera Ismail Khan (D.I. Khan) and Dir (Khyber Pakhtunkhwa, Pakistan). For checking the adequacy of the fitted models, analysis of variance technique is applied on the deviance residuals.

2. MATERIAL AND METHODS

Data
A record of 30 years (1970-1999) of daily rainfall data for D.I. Khan (31° 50', 70° 54') and 30 years (1970-1999) for Dir (35° 12', 71° 52') were obtained from the regional meteorological centre, Lahore, Pakistan. For this study, a rainy day is defined as being with at least 0.25 mm (0.01 inch) of rainfall.

It may be noted that the two sites have quite distinct weather pattern. This is obvious from Table 1 where some summary statistics for both the sites are given for quick comparison. This Table indicates that annual mean rainfall for Dir is more than 4 times greater than annual mean rainfall for D. I. Khan. In addition, some marked features of rainfall are also described for each site in the following sections.

<table>
<thead>
<tr>
<th>Site</th>
<th>Data</th>
<th>Annual Mean rain (mm)</th>
<th>Annual mean wet days</th>
<th>Max. yearly rain (mm)</th>
<th>Min. yearly rain (mm)</th>
<th>Wettest month</th>
<th>Driest month</th>
</tr>
</thead>
</table>
D.I. Khan

D.I. Khan is 340 Km south of Peshawar (the provincial capital of Khyber Pakhtunkhwa province). Some important statistics of the monthly rainfall are given in Table 2. For the given period, the average annual rainfall and average annual number of rainy days are 312.32 mm and 35.50, respectively. The wettest months are March, July and August; the driest months are October, November, December and January. The mean monthly rainfall and number of rainy days are presented in Table 3.

Table 2:
Some relevant Statistics of the monthly rainfall (D. I. Khan, 1970-99)

<table>
<thead>
<tr>
<th>Month</th>
<th>Amount of Rainfall (mm)</th>
<th>No. of wet days</th>
<th>Total No. of days</th>
<th>Mean rain per rainy day (mm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jan</td>
<td>323.13</td>
<td>89</td>
<td>930</td>
<td>3.63</td>
</tr>
<tr>
<td>Feb</td>
<td>626.11</td>
<td>108</td>
<td>840</td>
<td>5.79</td>
</tr>
<tr>
<td>Mar</td>
<td>1318.16</td>
<td>187</td>
<td>930</td>
<td>7.04</td>
</tr>
<tr>
<td>Apr</td>
<td>956.00</td>
<td>129</td>
<td>900</td>
<td>7.41</td>
</tr>
<tr>
<td>May</td>
<td>431.76</td>
<td>77</td>
<td>930</td>
<td>5.60</td>
</tr>
<tr>
<td>Jun</td>
<td>662.52</td>
<td>68</td>
<td>900</td>
<td>9.74</td>
</tr>
<tr>
<td>Jul</td>
<td>2234.17</td>
<td>132</td>
<td>930</td>
<td>16.92</td>
</tr>
<tr>
<td>Aug</td>
<td>1727.72</td>
<td>117</td>
<td>930</td>
<td>14.76</td>
</tr>
<tr>
<td>Sep</td>
<td>598.61</td>
<td>53</td>
<td>900</td>
<td>11.29</td>
</tr>
<tr>
<td>Oct</td>
<td>156.48</td>
<td>36</td>
<td>930</td>
<td>4.34</td>
</tr>
<tr>
<td>Nov</td>
<td>115.49</td>
<td>26</td>
<td>900</td>
<td>4.44</td>
</tr>
<tr>
<td>Dec</td>
<td>219.18</td>
<td>43</td>
<td>930</td>
<td>5.09</td>
</tr>
<tr>
<td>Total</td>
<td>9369.33</td>
<td>1065</td>
<td>10950</td>
<td>-</td>
</tr>
</tbody>
</table>

Table 3:
Mean and S.D of monthly rainfall amount and No. of rainy days (D. I. Khan, 1970-99)

<table>
<thead>
<tr>
<th>Month</th>
<th>Amount of rain</th>
<th>No. of rainy days</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>S.D.</td>
</tr>
<tr>
<td>Jan</td>
<td>10.77</td>
<td>12.63</td>
</tr>
<tr>
<td>Feb</td>
<td>20.87</td>
<td>33.30</td>
</tr>
<tr>
<td>Mar</td>
<td>43.93</td>
<td>38.27</td>
</tr>
<tr>
<td>Apr</td>
<td>31.86</td>
<td>30.95</td>
</tr>
<tr>
<td>May</td>
<td>14.39</td>
<td>15.29</td>
</tr>
<tr>
<td>Jun</td>
<td>22.08</td>
<td>23.11</td>
</tr>
<tr>
<td>Jul</td>
<td>74.47</td>
<td>98.52</td>
</tr>
<tr>
<td>Aug</td>
<td>57.59</td>
<td>44.66</td>
</tr>
<tr>
<td>Sep</td>
<td>19.95</td>
<td>28.57</td>
</tr>
<tr>
<td>Oct</td>
<td>5.21</td>
<td>9.55</td>
</tr>
<tr>
<td>Nov</td>
<td>3.84</td>
<td>8.48</td>
</tr>
<tr>
<td>Dec</td>
<td>7.30</td>
<td>12.43</td>
</tr>
<tr>
<td>Annual</td>
<td>312.31</td>
<td>163.91</td>
</tr>
</tbody>
</table>
The variation in monthly rainfall can be described by computing 20, 50 and 80 percentage points by ordering the total for each month and are given in Figure 1. It can be depicted from this Figure that the rainfall is nearly constant (and almost insignificant) from October to January. This Figure shows two regimes of rainfall for the concerned site. During the first regime, the rain very slowly starts from November and reaches to the third wettest month March, and then continues decreasing up to May. However, in the second regime, the rain abruptly increases from May to July and in a similar fashion continues decreasing abruptly up to November. In 50% of all years, rainfall is negligible for October, November and December.

![Figure 1: Percentage points of monthly rainfall, D. I. Khan (1970-99)](image)

The pattern of rainfall is bimodal. The variation is large in the wettest month, i.e. July where 20% of years have less than 20 mm and 20% more than 170 mm of rainfall. The annual amount of rainfall is shown in Figure 2. The minimum amount is 143.8 mm for 1979 and the maximum amount is 1026.67 mm for 1970.

![Figure 2: Annual rainfall for D. I. Khan (1970-99)](image)
Dir

Dir is about 225 km from Peshawar. It is a hill station with moderate summer. Table 4 comprises some relevant statistics of the monthly rainfall for Dir. For the period concerned, the average annual rainfall and average annual number of rainy days are 1391.67 mm and 102, respectively. The wettest months are February, March, April, July and August; whereas the driest month is November. The mean monthly rainfall and number of rainy days are illustrated in Table 5.

Table 4:
Some relevant Statistics of the monthly rainfall (Dir, 1970-99)

<table>
<thead>
<tr>
<th>Month</th>
<th>Amount of Rainfall (mm)</th>
<th>No. of wet days</th>
<th>Total No. of days</th>
<th>Mean rain per rainy day (mm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jan</td>
<td>3576.2</td>
<td>238</td>
<td>930</td>
<td>14.66</td>
</tr>
<tr>
<td>Feb</td>
<td>4777.3</td>
<td>262</td>
<td>840</td>
<td>18.23</td>
</tr>
<tr>
<td>Mar</td>
<td>7285.3</td>
<td>368</td>
<td>930</td>
<td>19.79</td>
</tr>
<tr>
<td>Apr</td>
<td>4793.9</td>
<td>348</td>
<td>900</td>
<td>13.77</td>
</tr>
<tr>
<td>May</td>
<td>2772.4</td>
<td>281</td>
<td>930</td>
<td>9.86</td>
</tr>
<tr>
<td>Jun</td>
<td>1546.0</td>
<td>227</td>
<td>900</td>
<td>6.81</td>
</tr>
<tr>
<td>Jul</td>
<td>4506.2</td>
<td>335</td>
<td>930</td>
<td>13.45</td>
</tr>
<tr>
<td>Aug</td>
<td>4468.5</td>
<td>328</td>
<td>930</td>
<td>13.62</td>
</tr>
<tr>
<td>Sep</td>
<td>2726.9</td>
<td>264</td>
<td>900</td>
<td>10.32</td>
</tr>
<tr>
<td>Oct</td>
<td>1805.8</td>
<td>161</td>
<td>930</td>
<td>11.21</td>
</tr>
<tr>
<td>Nov</td>
<td>1326.4</td>
<td>84</td>
<td>900</td>
<td>15.79</td>
</tr>
<tr>
<td>Dec</td>
<td>2165.2</td>
<td>162</td>
<td>930</td>
<td>13.36</td>
</tr>
<tr>
<td>Total</td>
<td>41750.1</td>
<td>3058</td>
<td>10950</td>
<td>-</td>
</tr>
</tbody>
</table>

Table 5:
Mean and S.D of monthly rainfall amount & No. of rainy days (Dir, 1970-99)

<table>
<thead>
<tr>
<th>Month</th>
<th>Amount of rain</th>
<th>No. of rainy days</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>S.D.</td>
</tr>
<tr>
<td>Jan</td>
<td>119.20</td>
<td>89.18</td>
</tr>
<tr>
<td>Feb</td>
<td>159.24</td>
<td>84.41</td>
</tr>
<tr>
<td>Mar</td>
<td>242.84</td>
<td>149.07</td>
</tr>
<tr>
<td>Apr</td>
<td>159.79</td>
<td>76.80</td>
</tr>
<tr>
<td>May</td>
<td>92.41</td>
<td>53.29</td>
</tr>
<tr>
<td>Jun</td>
<td>51.53</td>
<td>26.49</td>
</tr>
<tr>
<td>Jul</td>
<td>150.20</td>
<td>64.32</td>
</tr>
<tr>
<td>Aug</td>
<td>148.95</td>
<td>64.59</td>
</tr>
<tr>
<td>Sep</td>
<td>90.89</td>
<td>54.96</td>
</tr>
<tr>
<td>Oct</td>
<td>60.19</td>
<td>61.01</td>
</tr>
<tr>
<td>Nov</td>
<td>44.21</td>
<td>62.48</td>
</tr>
<tr>
<td>Dec</td>
<td>72.17</td>
<td>70.46</td>
</tr>
<tr>
<td>Annual</td>
<td>1391.67</td>
<td>304.92</td>
</tr>
</tbody>
</table>
Different quantiles (20\textsuperscript{th}, 50\textsuperscript{th} & 80\textsuperscript{th}) are demonstrated in Figure 3. To judge the variation in monthly rainfall from this Figure, it can be seen that rainfall can broadly be divided into two segments. In the first segment, the rain starts moderately from November and continues sharply increasing up to March and then sharply decreases until June. In the second segment, the rain abruptly increases from June to July and then starts decreasing rapidly till November is reached. It is obvious from this Figure that the rainfall is almost constant from July to August. The pattern of rainfall is bimodal. The variation is large in the wettest month, i.e. March, where 20\% of years have less than 118 mm and 20\% more than 360 mm of rainfall. The annual amount of rainfall is shown in Figure 4. The minimum amount is 882 mm for 1971 and the maximum amount is 2149 mm for 1986, respectively.

![Fig. 3: Percentage points of monthly rainfall (Dir, 1970-99)](image1)

![Fig. 4: Annual rainfall for Dir (1970-99)](image2)
Gamma Distribution:
The gamma distribution is quite often used for modeling the non-zero amount of rain (Stern & Coe, 1982, 1984, Siddiqui, 1992, Sharda & Das, 2005 and Hasan & Dunn, 2010). The p.d.f. of the gamma distribution is given by

$$F(Y) = \left(\frac{K}{\mu}\right)^K Y^{K-1} e^{-KY/\mu} / \Gamma(K), \quad K > 0, \quad 0 \leq Y \leq \infty.$$  \hspace{1cm} (1)

where $\mu$ is the mean rain per rainy day, $\Gamma$ is the gamma function and $K$ is the shape parameter. The coefficient of variation is given by $1/\sqrt{K}$.

The maximum likelihood (m.l.) method may be used for estimating the parameters of the gamma distribution. If $\mu$ (the m.l. estimate of which is $\bar{Y}$) is constant throughout, the m.l. estimate of $K$ can be obtained by solving the following equation.

$$\log(\hat{K}) - \psi(\hat{K}) = \log \bar{Y} - \left(\sum_{i=1}^{n} \log y_i / n\right)$$  \hspace{1cm} (2)

In equation (2), $\bar{Y}$ is the mean of the rainfall amount $y_1, y_2, ..., y_n$, and $\psi$ is the digamma function. This equation can be solved by using the Tables given by Pearson and Hartley (1970). However, in this study $K$ was estimated by solving the equation given by Mooley (1973) using Newton’s method.

It is pertinent to mention that in equation (2), the parameters of the $\mu$ & $K$ are assumed to be constant throughout; however, the assumption of constant parameters may not be very practical. In rainfall analysis, it is generally assumed that $K$ is constant, but the rainfall amount per rainy day may vary with time through the rainy season. One possibility of modeling this is to allow different values of $\mu$ for different months (Hussain, 1991); another possibility is to consider $\mu$ as continuously varying with time of year. Here, the second possibility is taken into account in the following section.

3. CONTINUOUS (TIME) MODELS FOR THE MEAN RAINFALL

It may be noted that fitting separate models for each month requires a large number of parameters. Yet, it is quite possible, specifically at the beginning and end of the rainy season, that the mean parameter (mean rain per rainy day) may change within a month.

To satisfy this possibility, continuous time models are considered, where $\mu(t)$ (mean rain per rainy day) depends on $t$ (day) through comparatively few parameters. Generally, the model is fitted to the $T$ days of the year from day $t_1$ to $t_T$, where $t = t_1, t_2, ..., t_T$.

It is worth mentioning here that the shape parameter, $K$, is usually assumed to be constant for the entire season, however, $\mu(t)$ is allowed to depend on the time of the year, which implies that $\mu$ varies continuously with time. In this study, the independent gamma distribution in which the amount of rain does not depend on the wet or dry state of the previous day, is fitted. Generalized linear models (Nelder & Wedderburn, 1972; McCullagh and Nelder, 1989) are used for fitting this distribution, as gamma distribution.
is a member of the exponential family. According to the components of generalized linear models the linear predictor, \( \eta \) (eta) is defined as

\[
\eta = \sum_{j=1}^{p} \beta_j X_j = \mathbf{X} \mathbf{\beta}
\]  

where the R.H.S. is the systematic component of the model and the covariates \( x_1, x_2, \ldots, x_p \) (with associated parameters \( \beta_j \)) produce the linear predictor \( \eta \).

It is to be noted that usually we have a link function \( g \) such that \( \eta_i = g(\mu_i) \). Here, a log link is taken in fitting a generalized model (since the mean rain per rainy day must be positive), so that linear predictor \( \eta = \log \mu(t) \) and hence \( \mu(t) = e^\eta > 0 \).

The linear predictor should be chosen carefully. In this regard, Stern and Coe (1982, 1984) have shown that Fourier series may be required to model a general (periodic) function \( \mu(t) \), with the number of harmonics selected for satisfactory results. However, a polynomial in \( t \) may also be used, alternatively (Hussain, 1991)

While fitting the model, the non-zero amount of rainfall are considered for all the 30 years (for D.I. Khan and Dir) as the dependent variable. The days of the rainy season or year (\( t_1 = \text{first day} \) to \( t_T = \text{last day} \)) on which rain occurred in each year are taken as the independent variable. The adequacy of the generalized linear models is checked by calculating the statistic, deviance, for each model. The deviance is defined as under:

\[
\text{Deviance} = -2 \left( \text{maximum log likelihood under } H_0 - \text{maximum log likelihood under } H_1 \right).
\]

In the sequel, the suitable model is chosen on the basis of reduction in deviance among several competing models involving Fourier harmonics or polynomial function of ‘days’.

Note that for generalized linear models with an unknown scale parameter, the deviance has an asymptotic distribution equal to the scale parameter times a chi-square distribution. This can be used for checking the significance of the reduction in deviance of the two models employing the ratio given below:

\[
\frac{\left( D_0 - D_1 \right) / (p-q)}{D_1 / (N-p)} = F_{p-q, N-p}
\]

where \( D_0 = \text{deviance of the model under } H_0; \ D_1 = \text{deviance of the model under } H_1 \)

\( N = \text{Total number of observations}; \ p = \text{No. of parameters to be estimated under } H_1 \)

\( q = \text{No. of parameters to be estimated under } H_0 \)

The results of the gamma distribution models fitted to the rainfall data of D.I. Khan and Dir are shown (separately) in the following.

**D.I. Khan.**

The results of fitting the gamma distribution model with \( \mu(t) = e^{\eta(t)} \) for linear predictor involving harmonic function in \( t \) (days) are given in Table 6. The linear predictor involving polynomial function in \( t \) was also fitted. However, only the
appropriate choice of linear predictor involving either polynomial function or harmonic function is considered here. It can be seen from Table 6 that Fourier series with three harmonics was found to be appropriate, yielding the lowest deviance. The corresponding coefficients along with Standard Error (S.E.) and P-values are also illustrated in this Table. Hence, using the estimates of the parameters in this Table, the mean amount of rain for each day of the year can be calculated.

### Table 6:
Deviance & Coefficients of Fourier Harmonics for the gamma distribution model for D.I. Khan

<table>
<thead>
<tr>
<th>Model</th>
<th>Deviance</th>
<th>D.F.</th>
<th>Reduction in Deviance</th>
<th>D.F.</th>
</tr>
</thead>
<tbody>
<tr>
<td>$a_0$ (Constant)</td>
<td>1827.39</td>
<td>1064</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>1 Harmonic</td>
<td>1635.18</td>
<td>1062</td>
<td>192.21</td>
<td>2</td>
</tr>
<tr>
<td>2 Harmonics</td>
<td>1605.09</td>
<td>1060</td>
<td>30.09</td>
<td>2</td>
</tr>
<tr>
<td>3 Harmonics</td>
<td>1591.47</td>
<td>1058</td>
<td>13.62</td>
<td>2</td>
</tr>
<tr>
<td>4 Harmonics</td>
<td>1583.20</td>
<td>1056</td>
<td>8.27</td>
<td>2</td>
</tr>
</tbody>
</table>

Estimates of coefficients for mean rain per rainy day

<table>
<thead>
<tr>
<th>Coefficient</th>
<th>Estimate</th>
<th>S.E.</th>
<th>P-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$a_0$ (Constant)</td>
<td>1.961922</td>
<td>0.04432886</td>
<td>0.000</td>
</tr>
<tr>
<td>$a_1$</td>
<td>-0.569331</td>
<td>0.06194812</td>
<td>0.000</td>
</tr>
<tr>
<td>$b_1$</td>
<td>-0.104289</td>
<td>0.06323701</td>
<td>0.099</td>
</tr>
<tr>
<td>$a_2$</td>
<td>0.019960</td>
<td>0.05837712</td>
<td>0.732</td>
</tr>
<tr>
<td>$b_2$</td>
<td>0.236143</td>
<td>0.06203092</td>
<td>0.000</td>
</tr>
<tr>
<td>$a_3$</td>
<td>-0.080386</td>
<td>0.05623436</td>
<td>0.153</td>
</tr>
<tr>
<td>$b_3$</td>
<td>-0.161627</td>
<td>0.05909634</td>
<td>0.006</td>
</tr>
</tbody>
</table>

For instance, the mean amount of rain, on 13th August (day 225 of the year) can be calculated as under.

The time of year, $t_i$, is taken as a day number from 1 to 365 and expressed in radians, considering a year as complete cycle from 0 to $2\pi$ by taking $t_i = (2\pi/365) \times i$, where $i= 1, 2, ..., 365$.

So, day 225, i.e. 13th August (in radians) = 0.017214 x 225. Hence,

$$t_{225} = 3.8732$$

Now, the estimates of parameters are substituted in the linear predictor $\eta$ (equation (3)) to get the mean amount of rain on day 225.

So,

$$\eta=1.96192-0.56933x\text{Cos}3.8732-0.10428x\text{Sin}3.8732+0.01996x\text{Cos}(2x3.8732)$$

$$+0.23614x\text{Sin}(2x3.8732)-0.08038x\text{Cos}(3x3.8732)-0.16162x\text{Sin}(3x3.8732)$$

Hence $\eta = 1.96192+0.82942 = 2.79134$. 

Therefore, \( e^{\eta} = e^{2.79134} = 16.302 \)

So, mean amount of rain on 13th August for D.I. Khan is 16.302 mm.

Before proceeding further (i.e. describing the continuous curve for mean rain per rainy day), it is pertinent to discuss the analysis of deviance and the shape parameter, \( k \).

**Analysis of Deviance**

The deviance for various models can be computed by fitting gamma distribution for modeling mean rain per rainy day. These include total deviance, between-day deviance and deviance due to different polynomial or harmonic terms. For a comprehensive analysis, all these deviances along with corresponding degrees of freedom (d.f.) can be represented by constructing Analysis of variance (ANOVA) Tables on the analogy of one-way ANOVA Tables used for normally distributed data.

These deviances are given in Table 7 where ‘between day’ and ‘within-day’ deviances are the main components of the total deviance. The total number of rainy days from January to December is 1065 in the 30 years. Forty eight days had no rain in any of the year, hence the d.f. for between day deviance is 316. The within-day deviance is that obtained by allowing \( \mu(t) \) to take a different value on each day. The between-day deviance is further divided into the contributions from the different harmonics and a residual term (which is the between-day effect not accounted for by the 1\textsuperscript{st}, 2\textsuperscript{nd} and 3\textsuperscript{rd} harmonics).

<table>
<thead>
<tr>
<th>Model</th>
<th>D.F.</th>
<th>Deviance</th>
<th>Mean Deviance</th>
<th>F. ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>Between Day</td>
<td>316</td>
<td>751.66</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>1 Harmonic</td>
<td>2</td>
<td>192.21</td>
<td>96.10</td>
<td>58.34</td>
</tr>
<tr>
<td>2 Harmonics</td>
<td>2</td>
<td>30.09</td>
<td>15.04</td>
<td>9.13</td>
</tr>
<tr>
<td>3 Harmonics</td>
<td>2</td>
<td>13.62</td>
<td>6.81</td>
<td>4.13</td>
</tr>
<tr>
<td>4 Harmonics</td>
<td>2</td>
<td>8.27</td>
<td>4.13</td>
<td>2.51</td>
</tr>
<tr>
<td>Residual</td>
<td>308</td>
<td>507.47</td>
<td>1.64</td>
<td>1.14</td>
</tr>
<tr>
<td>Within days</td>
<td>748</td>
<td>1075.73</td>
<td>1.43</td>
<td>-</td>
</tr>
<tr>
<td>Total</td>
<td>1064</td>
<td>1827.39</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

It is to be noted that deviance for, gamma distribution model, has an asymptotic distribution equal to the scale parameter times a chi-square distribution. Hence, the ratio of mean deviance will have an approximate F distribution. Note that the analysis of this Table also confirms that the Fourier series with three harmonics is appropriate for modeling mean rain per rainy day, at the 5\% level, for D.I. Khan.

In the continuation, an estimate of \( k \) is also required to complete the model for the distribution of rainfall amount. Keeping in view the current continuous time model, the question arises as how to estimate \( k \) when \( \mu \) is varying with \( t \) (day). In this situation, Coe and Stern (1982) suggested the maximum likelihood estimate of \( k \) as described below:
\[ \log(\hat{k}) - \psi(\hat{k}) = \text{Deviance (within-day) /}2N \] (5)

where \( N \) = total No. of rainy days in all the years. The within-day deviance is 1075.73 in Table 7, hence

\[ \log(\hat{k}) - \psi(\hat{k}) = \frac{1075.73}{2 \times 1064} = 0.5055 \]

It is to be noted that the R.H.S. of equation (5) is different from R.H.S. of equation (2). The equation (5) may be solved using Newton’s method given by Mooley (1973). We have applied Newton’s method to estimate \( k \) according to equation (2). The value of \( \hat{k} = 1.127 \). Thus we have a constant value of \( k \) throughout but a different value of \( \mu(t) \) for each day of the year.

The observed and fitted mean rain per rainy day is given in model in Figure 5. Here the curve fits the mean amount of rainfall data almost satisfactorily from January to December. The minimum value of mean rain per rainy day is 3.77 mm in the first week of January, whereas, its maximum value is 18.04 mm happens in the third week of July. The amount of rain for January and December is nearly constant, i.e. in the range of 3.76 mm to 4.64 mm. The curve in Figure 5 rises gradually from January to end of March and then remains somewhat constant up to May. The curve continues rising sharply from June to July and then decreases in a similar fashion upto middle of October, and finally from this point, it continues decreasing very slowly upto the last day of December, where the value 3.79 mm is attained. It is obvious that the main aim in any analysis is to summarize large values of data as concisely as possible, without discarding important information. Here the daily rainfall amount at D.I. Khan have been summarized by the equation of curve in Figure 5 and the value of \( k \).

![Fig. 5: Observed & fitted mean rainfall per rainy day for D.I.K. (1970-99)](image-url)
4. DIAGNOSTIC CHECKING OF THE MODEL

In generalized linear models, deviance is generally used to check the adequacy of a fitted model. The deviance for a gamma distribution is given as

\[ DV = 2 \sum \left[ -\log \left( \frac{y_i}{\hat{\mu}_i} \right) + \frac{(y_i - \hat{\mu}_i)^2}{\hat{\mu}_i} \right] \]

where DV stands for the deviance and sum extends over \( i = 1, 2, \ldots, N \);

The deviance is obtained by the contribution of quantity \( d_i \) for each unit, such that \( \sum d_i = DV \). Hence the deviance residual is defined as

\[ r_{DV} = \text{sgn} \left( y_i - \hat{\mu}_i \right) \times \sqrt{d_i} \]

where the sign (sgn) is that of \( (y_i - \hat{\mu}_i) \). So, we have a quantity that increases or decreases with \( y_i - \hat{\mu}_i \). It is obvious that \( \sum r_{DV}^2 = DV \). Thus, if there is any lack of fit in the Figure 5, this should be reflected in the deviance residuals from the fitted model with 3 harmonics.

It is significant to mention that if the model is not satisfactory, then this should be reflected in a tendency for the deviance residuals to vary systematically across the season. To this end, the deviance residuals were divided into twelve monthly groups, and an analysis of variance test is performed on them in Table 8. From Table 8, the hypothesis of equal monthly deviance residual means is accepted at the 5% level of significance. This confirms the view that the model fit is adequate.

Table 8:
Analysis of variance for the monthly deviance residual means for D.I. Khan (1970-99)

<table>
<thead>
<tr>
<th>Source of variation</th>
<th>D.F.</th>
<th>Sum of square</th>
<th>Mean squares</th>
<th>F-ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>Between Months</td>
<td>11</td>
<td>8.88</td>
<td>0.807</td>
<td>0.66</td>
</tr>
<tr>
<td>Within Months (Error)</td>
<td>1053</td>
<td>1286.12</td>
<td>1.221</td>
<td>( F_{0.05(11,1053)} = 1.83 )</td>
</tr>
<tr>
<td>Total</td>
<td>1064</td>
<td>1295.05</td>
<td>-</td>
<td>P-value=0.777</td>
</tr>
</tbody>
</table>

Dir

The relevant information of fitting the gamma distribution model for the mean amount of daily rainfall at Dir is presented in Table 9. It is evident from this Table that Fourier series with four harmonics is found to be satisfactory, yielding the lowest deviance. The estimates of the model coefficients along with S.E. and F-values are also given showing the significance of the parameters involved. The analysis of deviance for the gamma distribution model for Dir is given in Table 10, where ‘between’ and ‘within-day’ deviance are the major components of the total deviance. This Table depicts that the total number of rainy days from January to December is 3058 in the 30 years. One day had no rain in any of the year. So the d.f. for between-day deviance is 363. The analysis of deviance also confirms that Fourier series with four harmonics is required for modeling mean rain per rainy day for Dir at the 5% level.
Table 9:
Deviance & Coefficients of Fourier Harmonics
for gamma distribution model for Dir

<table>
<thead>
<tr>
<th>Model</th>
<th>Deviance</th>
<th>D.F.</th>
<th>Reduction in Deviance</th>
<th>D.F.</th>
</tr>
</thead>
<tbody>
<tr>
<td>$a_0$ (Constant)</td>
<td>4237.80</td>
<td>3057</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>1 Harmonic</td>
<td>4142.88</td>
<td>3055</td>
<td>94.92</td>
<td>2</td>
</tr>
<tr>
<td>2 Harmonics</td>
<td>4062.15</td>
<td>3053</td>
<td>80.73</td>
<td>2</td>
</tr>
<tr>
<td>3 Harmonics</td>
<td>4002.07</td>
<td>3051</td>
<td>60.08</td>
<td>2</td>
</tr>
<tr>
<td>4 Harmonics</td>
<td>3990.73</td>
<td>3049</td>
<td>11.34</td>
<td>2</td>
</tr>
</tbody>
</table>

Estimates of coefficients for mean rain per rainy day

<table>
<thead>
<tr>
<th>coefficient</th>
<th>Estimate</th>
<th>S.E.</th>
<th>P-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$a_0$ (Constant)</td>
<td>2.55314</td>
<td>0.022396</td>
<td>0.000</td>
</tr>
<tr>
<td>$a_1$</td>
<td>0.236064</td>
<td>0.032351</td>
<td>0.000</td>
</tr>
<tr>
<td>$b_1$</td>
<td>0.093473</td>
<td>0.030953</td>
<td>0.002</td>
</tr>
<tr>
<td>$a_2$</td>
<td>-0.077921</td>
<td>0.030534</td>
<td>0.011</td>
</tr>
<tr>
<td>$b_2$</td>
<td>0.188041</td>
<td>0.032594</td>
<td>0.000</td>
</tr>
<tr>
<td>$a_3$</td>
<td>-0.081901</td>
<td>0.031492</td>
<td>0.006</td>
</tr>
<tr>
<td>$b_3$</td>
<td>-0.190125</td>
<td>0.030550</td>
<td>0.000</td>
</tr>
<tr>
<td>$a_4$</td>
<td>-0.35334</td>
<td>0.030758</td>
<td>0.251</td>
</tr>
<tr>
<td>$b_4$</td>
<td>0.082611</td>
<td>0.029948</td>
<td>0.005</td>
</tr>
</tbody>
</table>

Table 10:
Analysis of deviance for modeling mean rainfall per rainy day at Dir, with a Fourier series

<table>
<thead>
<tr>
<th>Model</th>
<th>D.F.</th>
<th>Deviance</th>
<th>Mean Deviance</th>
<th>F. ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>Between Dy</td>
<td>363</td>
<td>714.58</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>1 Harmonic</td>
<td>2</td>
<td>94.92</td>
<td>47.46</td>
<td>36.04</td>
</tr>
<tr>
<td>2 Harmonics</td>
<td>2</td>
<td>80.73</td>
<td>40.365</td>
<td>30.65</td>
</tr>
<tr>
<td>3 Harmonics</td>
<td>2</td>
<td>60.08</td>
<td>30.04</td>
<td>22.81</td>
</tr>
<tr>
<td>4 Harmonics</td>
<td>2</td>
<td>11.34</td>
<td>5.67</td>
<td>4.31</td>
</tr>
<tr>
<td>Residual</td>
<td>355</td>
<td>467.51</td>
<td>1.317</td>
<td>1.01</td>
</tr>
<tr>
<td>Within days</td>
<td>2694</td>
<td>3523.22</td>
<td>1.308</td>
<td>-</td>
</tr>
<tr>
<td>Total</td>
<td>3057</td>
<td>4237.8</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

The m.l.e. of the shape parameter, $k$ using the within-day deviance was calculated according to the equation (5) as

$$\log (\hat{k}) - \psi (\hat{k}) = \frac{3523.22}{(2 \times 3058)} = 0.57606$$

The solution of the above equation using Newton’s method gives the value of $k = 1.001$. The observed and fitted mean rain per rainy day is shown in Figure 6. The curve in this Figure appears to pass (fit) through the observed points quite appropriately.
The minimum amount 6.98 mm of rain occurs in the first week of June, whereas, it maximum value 19.86 mm of the year occurs during the first week of March. It is interesting to note that the maximum mean amount of rain occurs in March instead of July and August (i.e. the monsoon season).

The curve in Figure 6 illustrates that the amount of rain first increases from January to beginning of March by reaching to its maximum mean value, and then decreases up to beginning of June where its minimum value of the year is reached. From here it again starts rising and reaches to its second peak by attaining the value of 14.98 mm at the beginning of August. From this point, the curve again declines upto end of September and then slowly rises up to middle of November; finally it decreases very slowly till the end of December where the value of 13.34 mm is attained.

The adequacy of the model in Figure 6 was checked by performing an analysis of variance test on deviance residuals on monthly basis (on the pattern of D.I. Khan rainfall data in Table 8). It was observed that the hypothesis of equal monthly deviance residual means is accepted at the 5% level of significance ($F_{\text{calculated}} = 0.97$ and $p$-value = 0.472). Thus it suggests that the model fits the data of Dir, satisfactorily.

5. CONCLUSION & RECOMMENDATION

In this study, the modeling approach to daily rainfall data is satisfactory. The gamma distributed amount of rain on rainy day lends itself to analysis using generalized linear models (with appropriate linear predictor) to reflect seasonal changes. It is interesting to note such changes and fluctuations are modeled using (in some cases) few parameters. The models fitted here are of much importance to agriculturists.

In this research, we concentrated on application of the modeling techniques to 30 years of daily rainfall data from D.I. Khan and Dir having quite distinct weather features. These sites may be classified into two categories according to the annual mean amount of rainfall. Here, D.I. Khan (with annual mean rainfall = 312 mm) may be classified as low
rainfall receiving sites, whereas, Dir (with annual mean rainfall = 1391.7 mm) may be classified as high rainfall receiving site.

The shape parameter, \( k \) is assumed to be constant throughout the season but \( \mu \) (mean amount of rainfall per rainy day) is allowed to take different value on each day \( t \) and was denoted by \( \mu (t) \). Under this assumption forms for \( \mu (t) \) were estimated by fitting the gamma distribution model to all non-zero observed rainfall amount. The models thus fitted for both D.I. Khan and Dir were found to be suitable for explaining the seasonal variation in the mean rainfall per rainy. An important conclusion is that the pattern of rainfall amount, throughout the year, is bimodal for both D.I. Khan and Dir. The first peak emerges in March, while the second peak emerges in July/August (i.e. during the monsoon season) as can be confirmed both from the graphs of percentage points of monthly rainfall and from the continuous curves for mean rain per rainy day for the two sites.

For checking the adequacy of the fitted models, analysis of variance test on deviance residuals was carried out. It was observed that the hypothesis of equal monthly deviance residual means is accepted. This indicates that the models fit the data of both the sites, adequately.

The information on certain aspects of rainfall like, the probability of rainfall and the distribution of rainfall amount throughout the year, the start of the rainy season and the risk of dry spell after sowing etc. are valuable to agriculturists.

The ability to model rainfall is important for agricultural decision making and crop management studies. We conceive that the models fitted in this study have much prospective in agricultural research and planning, and can be used to obtain important characteristic(s) relating to rainfall amount. The onus is on agricultural planners and researchers to specify the vital aspects of rainfall for the improvement of agriculture of area concerned.

REFERENCES


NON-PARAMETRIC AND PARAMETRIC DURATION ANALYSIS OF CORONARY ARTERY DISEASE WITH CHEST PAIN IN PAST

Mehwish Hussain¹, Nazeer Khan² and Mudassir Uddin³

¹ Department of Statistics, University of Karachi, Karachi, Pakistan
Email: mehwish.hussain@yahoo.com
² Department of Research, Dow University of Health Sciences, Karachi, Pakistan. Email: n.khan@duhs.edu.pk
³ Department of Statistics, University of Karachi, Karachi, Pakistan
Email: mudassir2000@hotmail.com

ABSTRACT

This paper focuses on non-parametric and parametric survival models that best describe the distribution of duration of coronary artery disease (CAD) patients who suffered chest pain in past. The data contained information of 17,232 Saudi population sampled from a cross-sectional study; mainly investigating prevalence of CAD with certain risk factors and clinical findings for CAD diagnosis. Descriptive analysis revealed 86% of the 1,036 CAD patients had chest pain in past. Data were then converted from snapshot to time-dependent for advanced duration analysis.

For non-parametric duration analysis, Kaplan-Meier and Log-Rank analyses were performed to estimate the overall duration and stratified duration with respect to gender, marital status, education level etc. Stepwise Cox proportion hazard regression model was run to predict the duration in presence of significant factors and covariates. The proportional hazard function was first checked before constructing Cox regression model while graphing \(-\ln(\ln(\text{Survival Probability}))\) against different factors.

Parametric survival analysis was run while checking the distribution of CAD duration following either Weibull, exponential, log-logistic or generalized gamma distribution. The Accelerated Failure Time (AFT) model with minimum AIC value was considered good-fitted model for the duration of CAD from chest pain.

KEYWORDS

AFT Model, CAD, Duration Analysis, Model Selection, Proportional Hazard Assumption, Stepwise Regression

INTRODUCTION

Duration – a word defines as “the length of the time that something lasts or continues”¹. The analysis of such length of time, so called duration, is done by a statistical technique, “Survival Analysis”. With the name of this technique, a research based on long length of time, concurrently come in mind. Due to this reason researcher

Non-Parametric and Parametric Duration Analysis of Coronary Artery Disease (CAD) seeks to apply this technique for studying duration based on longitudinal study (Morita et al., 1989).

In medical settings, the analysis based on prospective study is quite cumbersome, demands lot of time, finance, keen supervision and effective manpower. Thus, many data produced in such setting would have been done either retrospectively or based on cross-sectional (Keiding et al., 2002).

During the last decade of the previous century, many statisticians started working on the aspect of deriving the literature for applying the same technique while using current or retrospective duration. The theoretical justifications have been presented in different papers (Keiding et al., 2005, Keiding et al., 2002, Keiding, 2006, Joffe et al., 2005).

This paper will explore the parametric and non-parametric duration analyses for coronary artery disease (CAD). Researches on CAD usually surrounds on prevalence findings or clinical diagnosis of related morphology (Hemingway et al., 2008, Bhopal, 2000). Though, there are certain studies based on the survival of CAD patients (Romero-Corral et al., 2006). Only single study was found by the author elucidating duration of myocardial infarction release to the hospitalized patients using longitudinal study design (Mingala and Estolano, 2007).

This study based on the duration of patient’s diagnosis of CAD having chest pain in past. The descriptive and basic non-parametric analysis has been presented earlier (Hussain and Khan, 2011).

THE DATA

The data was collected based on cross sectional study from 1995-2000, mainly finding prevalence of CAD in Kingdom of Saudi Arabia. Several probable risk factors have been also included for find any associations with the presence of CAD. Clinical diagnosis and parameters of CAD were also induced. The duration of having chest pain in past was asked by the patients verbally. For details, please see Al-Nozha et al., 2004.

INITIAL FINDING

A total of 17,232 subjects were participated in this study. It was found that 5.5% of the Saudi population prevalent with coronary artery disease (CAD). Significantly associated risk factors were Age, sex, body mass index (BMI), waist circumference (WC), systolic & diastolic blood pressures (SBP & DBP), smoking status, fasting blood sugar (FBS), serum cholesterol (S. Chol.), serum triglycerides (S.TG) and high density lipoprotein (HDL) (Al-Nozha et al., 2004).

CHEST PAIN DURATION FINDING

Among 1,036 CAD patients, 884 (85.3%) suffered with cardiac chest pain in past. Median duration was 5 years. The patients bearing chest pain since 17 and at least 25 years were more prone to have CAD. Males had more duration from chest pain to the diagnosis of CAD than females (Hussain and Khan, 2011). These are shown in figure 1 below.
Fig. 1: Kaplan Meier Survival Graph with respect to Gender, showing duration of both genders from chest to CAD.

STATISTICAL TECHNIQUE

Assuming the duration is positive and counted in whole number of integers. The distribution of duration is also assumed to be heavily tailed positively skewed. This implies that patients bearing chest pain with short period have high rate of not suffering with CAD and the patients with long duration of chest pain, are more likely to have CAD immediately. Thus, duration from chest pain to CAD is a non-increasing function.

Thus,

\[ t_{(1)} < t_{(2)} < t_{(3)} \ldots < t_{(k)}, \text{ then } S(t) = P(T > t) \]  

NON-PARAMETRIC MODEL

The proportional hazard model for the duration of CAD, also known as Cox model in the literature, assumes that

\[ \lambda(t/d) = \lambda_0(t) \exp(\sum_j d_j \beta_j) \]  

where \( \lambda(t/d) \) is the hazard of suffering with CAD at time ‘t’ given duration values \( d = (d_1, d_2, d_3, \ldots, d_k) \) and \( \lambda_0(t) \) is an arbitrarily baseline hazard function. The usual estimates of the parameter \( \beta = (\beta_1, \beta_2, \beta_3, \ldots, \beta_k)^T \) in the proportional hazard model without the specifications of \( \lambda_0(t) \) through maximization of the partial likelihood is

\[ L(\beta) = \prod_{r \in C} \frac{\exp(\beta^T x^r)}{\sum_{\tilde{x} \in R_r} \exp(\beta^T \tilde{x})} \]  

In expression 3, C is the set of indices of the diagnosis of CAD to patients, R, is the set of indices of the individuals at risk at time \( t_r \to 0 \) and \( j_r \) is the index of the CAD diagnosis at time \( t_r \).

**Variable Selection Criteria:**

There are several methods described for deciding which covariates should be used in proportional hazard models. These methods are the same methods used in ordinary least square regression. Indeed, the procedure of applying these variable selection techniques in PH model is same as that applying in logistic regression. Lasso, LAR, best subset, forward, backward and stepwise selection criteria are the most commonly used method (Tibshirani, 1997). The discussion on which method should be used in this particular dataset is put for the next paper. Here, stepwise selection method will be used while taking probability of entry = 0.05 and probability of stay = 0.10 (Collett, 2003).

The procedure of stepwise selection method will be: All the significant variables as described in Nozha et al. (2003) will be chosen for the model. Initially, the model initiated with a constant (no covariate). In the second stage, the most significant variables (having \( p \) value < 0.05) will be included in the model. The third stage will include the other most significant variable while checking the significance of already present first variable. If first included variable’s \( p \) value will become more than 0.10 with the inclusion of second variable, it will be excluded from the model else remain stayed. The second and third steps will be repeated until all the significant variables included in the model and insignificant will not be the part of model.

**PARAMETRIC MODEL**

**Distributional Assumptions:**

1. \( \hat{S}(t) = P(T > t) = \int_t^\infty f(d) \, d(d) \)
2. \( f(t) = h(t)\hat{S}(t) \)

where \( \hat{S}(t) = \exp(\int_0^t h(d) \, d(d)) \)
(Kleinbaum and Klein, 2005)

**Exponential Distribution:**

Let \( f(t) \) followed exponential distribution of the form:

\[ f(t) = \lambda e^{-\lambda t} \]

Then Survival Function will be \( S(t) = e^{-\lambda t} \) and hazard function equals \( H(t) = \lambda \).

**Gamma Distribution:**

If \( f(t) \) followed gamma distribution of the form:

\[ f(t) = \lambda^\alpha (\alpha)^{-1} t^{\alpha-1} e^{-\lambda t} \]

where \( \alpha > 0 \) is shape parameter and \( \lambda > 0 \) is scale parameter.

Then Survival Function will be \( S(t) = 1 - \int_0^t \lambda^\alpha (\alpha)^{-1} e^{-\lambda x} \, dx \) and hazard function equals \( H(t) = \frac{f(t)}{S(t)} \).
**Weibull Distribution:**

The Weibull distribution with the following density function

\[ f(t) = \lambda pt^{p-1}e^{-\lambda t^p} \]

bears survival function \( S(t) = e^{-\lambda t^p} \) and hazard function \( H(t) = \lambda pt^{p-1} \) with \( \lambda > 0 \) is scale parameter and \( p > 0 \) is shape parameter.

**Log-Logistic Distribution:**

With \( \alpha \) (as a scale parameter) > 0 and \( \beta \) (as a shape parameter) > 0, the density function, survival function and hazard function of log-logistic distribution are respectively given below:

\[ f(t) = \left( \frac{\beta}{\alpha} \right) \left( \frac{t}{\alpha} \right)^{\beta-1} \left[ 1 + \left( \frac{t}{\alpha} \right)^{\beta} \right]^{-2}, \]
\[ S(t) = \frac{1}{1 + \left( \frac{t}{\alpha} \right)^{\beta}}, \]
\[ H(t) = \left( \frac{\beta}{\alpha} \right) \left( \frac{t}{\alpha} \right)^{\beta-1} \frac{1}{1 + \left( \frac{t}{\alpha} \right)^{\beta}} \]

**Log-Normal Distribution:**

The positive skewed log-normal distribution has density function:

\[ f(x; \mu, \sigma) = \frac{1}{x\sigma\sqrt{2\pi}} e^{-\frac{(\ln x - \mu)^2}{2\sigma^2}} \]

\( S(t) = 1 - \phi \left( \frac{\ln(t) - \mu}{\sigma} \right) \) and \( H(t) = \frac{f(t)}{S(t)} \) are survival and hazard functions of log-normal distribution respectively. Here \( \mu \): location parameter and \( \sigma \): scale parameter.

**Gompertz Distribution:**

Gompertz is nothing but, a log-Weibull distribution, having following density function

\[ f(x; b, \eta) = be^{-bx} e^{-\eta x} e^{-bx} \left[ 1 + \eta (1 - e^{-bx}) \right] \]

where \( b > 0 \) is the scale and \( \eta > 0 \) is the shape parameter of the shifted Gompertz distribution.

\[ S(t) = \frac{f(t)}{H(t)} \] and \( H(t) = \exp(\alpha t) \)

**Goodness of Fit Criteria for AFT models:**

Among different goodness of fit testing methods, the most common used method was chosen i.e. the model with minimum AIC value will be considered as good fitted AFT model for the duration of CAD. The AIC value is defined as:

\[ \text{AIC} = -2 \times (\log \text{- likelihood}) + 2(p + 1 + s), \]

where \( p \) denotes the number of covariates in the model not including the constant term, \( s = 0 \) for the exponential model and \( s = 1 \) for the Weibull and log-logistic models (Hosmer et al., 2008).
DATA APPLICATION FINDINGS

Non-Parametric Proportional Hazard Model:

The proportional hazard (PH) assumption was checked by plotting $-\ln(-\ln(\text{Survival Probability}))$ by the comparison of different factors. All the variables fulfilled the PH assumptions portraying no interaction between different categories. All the significant variables were put in the model and stepwise Cox Regression Model was run.

Stepwise Cox regression model portrayed age, exercise, triglyceride nitrates (TGN), body mass index (BMI) and high density level (HDL) were the significant predictors that best describe the distribution of duration from chest pain to CAD with the estimated model:

$$\log_e (D) = 0.009 \text{ age} + 0.180 \text{ exercise} + 0.099 \text{ TGN} - 0.103 \text{ BMI} + 0.036 \text{ HDL}$$

($<0.0001$) ($0.045$) ($<0.0001$) ($<0.0001$) ($0.06$)

†P Values in parentheses

Parametric Models:

Table 1 shows the summarized table for the output obtained by the four distributions. Though, gamma and gompertz distributions were also applied, however these models were not estimated by the software STATA v. 11.0.

As per criteria of model fitting defined earlier, log-normal distribution has lowest AIC value, thus duration from chest pain to coronary artery disease follows log-normal distribution.

Figure 1 displayed the exploratory analyses for these distributions. All the models seemed to be good fitted. Though, exponential curve initiated with low survival (around 80%) and convexed quick, leaned to horizontal axis soon (at age about 25). Weibull initiated with a survival of about 90% and inclined to abscissa sooner than exponential (at age around 20). Log-normal and log-logistic showed initial survival more than that of weibull and exponential. However, log-logistic did not bent closely to x-axis even after the range of the data duration finished, implying presence of survival with long duration. On the other hand, log-normal displayed a smooth curve, fitted linearly to the original survival of duration of data.

Table 1:

<table>
<thead>
<tr>
<th>Distribution</th>
<th>Model</th>
<th>P Value</th>
<th>Log-Likelihood Value</th>
<th>AIC Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exponential</td>
<td>1.1673 Gender</td>
<td>0.0006</td>
<td>-2148.35</td>
<td>4300.702</td>
</tr>
<tr>
<td>Weibull</td>
<td>1.2107 Gender + 0.2173 ln(p) + 1.2428 p + 0.8046 p^{-1}</td>
<td>0.0006</td>
<td>-2098.55</td>
<td>4203.098</td>
</tr>
<tr>
<td>Log-Normal</td>
<td>1.5815 – 0.1138 Gender – 0.0833 ln (σ) + 0.9200 σ</td>
<td>0.0149</td>
<td>-2019.56</td>
<td>4045.121</td>
</tr>
<tr>
<td>Log-Logistic</td>
<td>1.5921 – 0.1346 Gender – 0.5994 ln (γ) + 0.5492 γ</td>
<td>0.0059</td>
<td>-2059.27</td>
<td>4124.547</td>
</tr>
</tbody>
</table>
CONCLUSION

Non-parametrically age, exercise, triglyceride nitrates (TGN), body mass index (BMI) and high density level (HDL) were the significant predictors that best describe the distribution of duration from chest pain to CAD. Log-Normal Distribution is the good-fitted distribution for describing CAD duration. This finding is different from earlier work on similar aspect.

REFERENCES


A BRIEF REVIEW OF GEOINFORMATICS ANALYSIS ON POVERTY DATA IN INDONESIA

Asep Saefuddin
Department of Statistics, Faculty of Mathematics and Natural Sciences
Bogor Agricultural University, Indonesia. Email: asaefuddin@gmail.com

ABSTRACT

A research group on geoinformatics was built in 2010 at The Department of Statistics of Bogor Agricultural University, Indonesia, to implement statistical analysis on spatial data. The initial step was to compile some statistical methods related on geographical regression of the simple approach and the complex ones. The methods were implemented to analyse the poverty data in Indonesia. Outcome of selected models was on poverty indicators in a district, such as percentage, expenditure per capita. The outcome was a priory influenced by some regional factors, i.e. local government policy, agro-climate typology, as well as local socio-culture. Therefore, the type of data may produce problems of outliers, outcome dependency, and non-stationarity. Classical approaches assuming there is no effect of the regional differences are not valid any more. For the first phase, the group then implemented methods related on GWR (Geographically Weighted Regression) including simultaneously or conditionally autoregressive models. To obtain firm statistical conclusion, the selected models were contrasted to the ordinary regression. The result indicated that district or regions affected on the poverty level. Hence, spatial factors cannot be neglected in analysing poverty in Indonesia. Additionally, geographical regression performed better than the ordinary models.

KEYWORDS
geographicall regression; regional factors; non-stationarity; spatial data.

1. INTRODUCTION

Many statistical approaches in geo-related sciences (geo-sciences) including social and environmental sciences have been developing rapidly due to the need in analyzing spatial data. The data found in these areas are usually based on observational study affected by geographical factors of both natural and non-natural effects. The natural factors may come from the type of agro-ecological typology, while the non-natural ones are due to the local governmental policies and socio-cultural aspects which may vary from place to place. In addition, the usual object of study is spatial unit comprising individuals or household in specific places. These conditions will inevitably introduce the problems of spatial heterogeneity, spatial non-stationarity, and spatial dependency in which the data are obtained. Spatial heterogeneity violates the assumption of variance homogeneity of model residuals (see for example in Getis and Ord, 1992) termed as error propagation by Haining (2004; Section 4.1.3). Spatial non-stationarity causes non-representativeness of global regression models explained nicely by Fotheringham,
Brunsdon and Charlton (2002). Spatial dependency violates the assumption of residual independence of regression models. In brief, standard statistical approaches are not appropriate any more and need some adjustment in analyzing spatial data. Otherwise, the models will produce misleading conclusion and recommendation. Anselin and Sergio (2010) provide a comprehensive review of literature on spatial statistics encompassing spatial analysis, pattern analysis, local statistics, and application. Spatial data analysis is commonly implemented in the regional science which includes economic, epidemiology, public health, and environmental sciences. Haining (2004) provide an excellent book related on the theory and practice of spatial data analysis.

Indonesia lies between latitudes 11°S and 6°N with longitudes 95°E and 141°E, a large country with 497 regencies and administrative cities spread out in 33 provinces. The variability of spatial data cannot be avoided due to the natural factors as well as non-natural ones including the local governments. The local governments in Indonesia since 2000, according the national regulation, have obtained their autonomy to govern the local programs except for the international policies. Therefore, the local policies will inevitably induce variabilities of some socio-economic variables. In brief, applying adequate methods related to spatial data is getting crucial in the country. Therefore the Department of Statistics of Bogor Agricultural University has initiated to run research and development in spatial statistics under a group named Geoinformatics. The objectives are to compile, to develop, and to implement spatial statistics in poverty data in Indonesia. Poverty data is an appropriate example for applying spatial statistics as the poverty is affected by the natural and non-natural factors. The term of geoinformatics is interchangably used with spatial statistics or geo-statistics. However, the GIS (Geographic Information System) part is not elaborated deeply in the paper making a more complete definition of geoinformatics according to Patil (2005) is not fully followed. The approach is concentrated on statistical modeling and analysis.

Poverty is one of the many complicated problems in developing countries caused by many factors which may be related each others. Spatial statistics is a powerful method to analyze factors causing poverty. First step to recognize the poverty of data is to implement EDA (exploratory data analysis) including descriptive statistics. In addition, EDA following Good (1983) may detect the pattern of data, identify interesting behaviour of data, recognize some errors and outliers, and formulate some hypotheses. In relation to spatial data, the EDA may be termed as ESDA (exploratory spatial data analysis) in which ‘location’ factor is included in the analysis (Haining, 2004). The ESDA is then very useful to firstly recognized the characteristics of data in a specific regions or factors related to the location (space). However, more detail analysis related to statistical relationship, cause-effect models, and prediction are not facilitated in the EDA or ESDA. Davenport and Harris (2008) mention that the more complicated the problems the more statistical intelligent is needed, including statistical analysis and modeling. However, descriptive statistics in EDA/ESDA is still a prerequisite for advance statistics.

2. PRELIMINARY RESULT

In this section preliminary results related to standard spatial statistics are described, i.e. descriptive statistics, geographically regression and autoregressive models for spatial data. The rest of the approaches are covered partially in the section of model
development. The more complex models related to overdispersion, random effect models, and spatial generalized linear models are not elaborated yet in this paper. Hence the paper concentrates on simple approaches but very important to consider.

2.1 Descriptive Statistics

Descriptive statistics found that the poverty in Indonesia was related to local government, i.e. country, district and regency, geographical relative position, shoreline, relative position from the forest, income and expenditure, housing condition, health and education facility, type of drinking water, accessibility, and many other factors related to socio-economic condition (Saefuddin et al 2011). In terms of covariates, Saefuddin et al (2011) also found that poverty in Indonesia is a function of main family income, transportation infrastructure, energy used for cooking, and local basic need facilities especially for education, health and economic activities. Unfortunately the ESDA cannot provide deeper information related to correlation between variables, estimation, prediction, and other advanced statistics. The group hence implement some statistical method related to spatial data, i.e. Spatial Regression, Geographically Weighted Regression (GWR), Spatial Autoregressive Regression (SAR), Conditional Autoregressive Regression (CAR), SAR Poisson, Spatial Empirical Best Linear Unbiased Prediction (SEBLUP), Spatial Bayesian Model of Small Area Estimation (SB-SAE), Hot-spot in SAE, GWR-Poisson Models. Some of the approached have been reported in MSc thesis and journal publications, while the rests are still in the research process for MSc and PhD levels. However, in this paper the author concentrated only in geographically weighted regression and autoregressive models which have been implemented by the group to analyze poverty in Indonesia.

2.2 Geographically Weighted Regression

In regional based modelling, the ordinary regression (OR) models provides global parameter estimates ($\beta$) assuming there is no spatial variation over regions, i.e. spatial stationarity. Unfortuantely, if the assumption is not fulfilled by empirical data, i.e. spatial non-stationarity, the global estimates will not be reliable. Theoretically the paremeter estimates are unbiased and applicable universally to all areas of study. Also, the models do not provide specific parameter estimates depending on the locations. However, geographically Weighted Regression (GWR) is relatively a new alternative approach to construct model on spatial non-stationarity condition (Brunsdon, Fotheringham and Charlton 1996, 1999; Fotheringham, Brunsdon and Charlton 2002). This approach is applied to test the structural stability of the predictors over region because non-stationarity data can compromise global result (Lambert et al 2006). By using GWR, it allows to estimate local models and produce local statistics specifically for every region. Fotheringham, Brunsdon and Charlton (2002) provide detail comparison between global and local estimates which is re-presented in Table 1.
A brief review of geoinformatics analysis on poverty data in Indonesia

Table 1: Differences between local and global statistics

<table>
<thead>
<tr>
<th>Global</th>
<th>Local</th>
</tr>
</thead>
<tbody>
<tr>
<td>Summarize data for whole region</td>
<td>Local disaggregation of global statistics</td>
</tr>
<tr>
<td>Single-valued statistic</td>
<td>Multi-values statistic</td>
</tr>
<tr>
<td>Non-mappable</td>
<td>Mappable</td>
</tr>
<tr>
<td>GIS unfriendly</td>
<td>GIS friendly</td>
</tr>
<tr>
<td>Aspatial or spatially limited</td>
<td>Spatial</td>
</tr>
<tr>
<td>Emphasize similarities across space</td>
<td>Emphasize differences across space</td>
</tr>
<tr>
<td>Search for regularities or ‘laws’</td>
<td>Search for exceptions or local ‘hot-spots’</td>
</tr>
</tbody>
</table>

(Source: Fotheringham, Brunsdon and Charlton, 2002, pp. 6)

Suppose, response variable of $n$ observations of is formed as a linear function of $p$ predictors. In OR, the model will be generally expressed as

$$y_i = \beta_0 + \sum_{k=1}^{p} \beta_k x_{ik} + \varepsilon_i$$

where $y_i$ is the outcome of $i$-th observation, $x_{ik}$ is $i$-th observation of $k$-th predictors, $\beta_k$ is $k$-th regression parameters ($k = 1,2,\ldots,p$) and $\varepsilon_i$ is error term that follows normal distribution with zero mean and known standard deviation and $i = 1,2,\ldots,n$. In matrix notation, if $X$ is an $n \times (p+1)$ matrix of predictors, $Y$ is an $n \times 1$ column vector of responses, $\beta$ is an $(p+1) \times 1$ column vector of parameter and $e$ is column vector of error term, Equation (1) can be written as

$$y = X\beta + e$$

OLR model is applied to all regions universally. However, in spatial case, it may be reasonable to assume that the effect of a predictor is conditional upon localized, unobserved factors and social networks (Lambert et al 2006). GWR method generates local model for every region, hence spatial effect could be accommodated better than a global model does. GWR is an extension of weighted regression in which the weighting function is based on the relative position among regions. The higher weight is assigned to nearby observation then it will decrease continuously when the distance farther away.

Equation (1) or (2) is considered as global model. Then if linear GWR model is applied, each \(i\)-th region of interest will have its own local model expressed as

\[
y_i = \beta_0(u_i, v_i) + \sum_{k=1}^{p} \beta_k(u_i, v_i)x_{ik} + \varepsilon_i
\]

where \((u_i, v_i)\) is location of \(i\)-th observation, or on the other word the centroid of \(i\)-th region (Fotheringham, Brunsdon and Charlton 2002). Suppose the diagonal elements of an \(n \times n\) matrix \(W(u_i, v_i)\) consist of the weight at region \((u_i, v_i)\). In matrix notation, for given \(y_i\), then

\[
y_i = X_i(X^TW(u_i, v_i)X)^{-1}X^TW(u_i, v_i)y
\]

The \(W(u_i, v_i)\) would be vary as \((u_i, v_i)\) is changed depending on its location (Leung et al 2000a). When all elements of \(W(u_i, v_i)\) remains constant over region, the ordinary weighted regression is held. However, in GWR, the weight at region \(i\)-th correspond to \(j\)-th region, \(w_j(i)\), is a function of distance of two regions, \(d_{ij} (j = 1, 2, \ldots, n)\).

To define weighting function, there is two common method can be applied, i.e. Gaussian function and bisquare function (Brunsdon et al 1996; Leung et al 2000; Fotheringham et al 2002). Related to the process of choosing the weighting function, it is very important to predetermine an optimum bandwidth. This could be done by minimizing either cross-validation score or Akaike Information Criteria (Forteringham et al 2002).

Regarding to residual sum of squares (RSS), GWR almost performs better than OLR does. However, GWR model needs statistical test to examine whether difference between RSS of GWR and RSS of OLR model, commonly named as GWR improvement, is significant or not. The test is actually to measure goodness-of-fit of GWR model compared to OLR model, that is to test null hypothesis that GWR and OLR model describe data variability equally well against alternative hypothesis that GWR model has better fit than OLR model does. There are several statistical procedures could be employed to perform the examination, i.e. ANOVA or \(F\) test (Brunsdon, Forteringham, and Charlton 1999; Forteringham, Brunsdon and Charlton 2002), \(F1\) and \(F2\) test (Leung, Mei, and Zhang 2000a). If GWR improvement is statistically significant, test for parameters variability is needed. The authors also proposed \(F3\) test which takes place in diagnosing whether the set of parameters tend to be constant or to be vary over region. Alternatively, Monte Carlo significance test procedure (Hope 1968) can also be used for such diagnostics.

In Indonesia, the application of the GWR is very limited. However, GWR has been getting implemented and developed, especially by the research group on geoinformatics. This paper presents brief review of GWR application examples in Indonesia which have been presented by the group, i.e. Saefuddin, Setiabudi and Achsani (2011) and Rahmawati (2010), to analyze poverty.

Saefuddin et al (2011) model poverty as a linear function of Human Development Index (HDI) of 116 districts and administrative cities in Java, Indonesia for the year of
A brief review of geoinformatics analysis on poverty data in Indonesia

2008 using data from The National Team for Accelerating Poverty Reduction, Office of Vice President of Republic of Indonesia (2010). Due to the existing Java condition, i.e. (1) differences of provincial and district governments, (2) distance of regions to central government, (3) regional autonomy, and (4) differences of agro-ecosystem or climate. Hence, it is reasonable to implement the GWR to analyze relationship between poverty and HDI rather than the OLR.

OLR model of poverty is firstly estimated follows simple linear form of

\[ \text{POV} = \beta_0 + \beta_1 \times \text{HDI} \]  

which is describing the relationship between the numbers of poverty as a percentage of population numbers (POV) and Human Development Index (HDI). The result of analysis is presented in Table 2 with \(RSS=3004.52, AICc=712.95, AIC=708.74\) and \(R^2=47.56\%\). All corresponding p-values of coefficients in this model indicate that intercept and HDI significantly affect poverty.

Table 2: Parameter estimates of poverty model using OLR

| Variable | \(\hat{\beta}\) | \(S(\hat{\beta})\) | \(t\) | \(Pr(>|t|)\) |
|----------|----------------|------------------|------|-------------|
| Intercept | 104.99 | 8.78 | 11.95 | <.0001 |
| HDI | -1.27 | 0.12 | -10.17 | <.0001 |

(Source : Saefuddin, Setiabudi and Achsani, 2011, pp. 281)

Since OLR model seems unsatisfied regarding to \(RSS\) and \(R^2\), then local models for every point in the region of interest were estimated using GWR. In selecting bandwidth of weighting function, all possible method combinations are utilized and then compared. Finally, the combination of Gaussian function and CV score approach produced an optimum bandwidth. Once, linear GWR was applied using Gaussian weighting function and the optimum bandwidth \(RSS\) decreased to 1376.66, \(R^2\) increased to 76\%, and \(AICc\) and \(AIC\) decrease to 661.38 and 633.92 respectively.

To test GWR improvement, three methods of goodness-of-fits test are performed. Table 3 summaries results of ANOVA or \(F\), \(F1\) and \(F2\) test. According to \(p\)-values, which were all smaller than 0.05, it could be concluded that that GWR model performed better fits than OLR model did. On the other word, GWR improvement was significant.

Table 3: Summary of GWR improvement test for poverty model

<table>
<thead>
<tr>
<th>Test</th>
<th>(df_1)</th>
<th>(df_2)</th>
<th>(F)-statistic</th>
<th>(p)-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>(F) or ANOVA</td>
<td>114.00</td>
<td>98.24</td>
<td>2.18</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>(F1)</td>
<td>100.24</td>
<td>114.00</td>
<td>0.57</td>
<td>0.0021</td>
</tr>
<tr>
<td>(F2)</td>
<td>34.31</td>
<td>114.00</td>
<td>2.77</td>
<td>&lt;.0001</td>
</tr>
</tbody>
</table>

(Source : Saefuddin et al 2011, pp. 282)

Similar to Saefuddin et al (2011), Rahmawati (2010) previously compared OLR and GWR model of per-capita expenditure per month (EXPE) of 35 villages in regency of
Jember, East Java, Indonesia as a function of socio-economic indicators for the year of 2008. Initially, there were 18 predictors selected for the model. However, only three variables were included on the model as a result of variable selection. Those were distance from village to capital of regency (DIST), number of health facility in every 1000 inhabitant (HFAC) and number of beneficiary family of ASKESKIN or health assurance for poor community as percentage of number of total family (ASKES). The general form of global model is

\[ \text{EXPE} = \beta_0 + \beta_1 \ast \text{DIST} + \beta_2 \ast \text{HFAC} + \beta_3 \ast \text{ASKES} \]  

(6)

| Variable  | \( \hat{\beta} \) | \( S(\hat{\beta}) \) | \( t \) | \( Pr(>|t|) \) |
|-----------|-----------------|-----------------|-------|---------------|
| Intercept | 147313.4        | 80592.8         | 1.83  | 0.077         |
| DIST      | -2978.2         | 951.8           | -3.13 | 0.004         |
| HFAC      | 175907.0        | 37931.6         | 4.64  | 0.000         |
| ASKES     | -2015.1         | 739.2           | -2.73 | 0.010         |

(Source : Rahmawati, 2010, pp. 19)

Fitting an OLR model in the form of equation (6) provides the results in Table 4. According to \( p \)-value on the table, predictors were all significantly affect per-capita expenditure. However, the model above is not quite feasible for use with the \( R^2=64.7\% \) and \( RSS=16.69 \times 10^{10} \).

To improve model accuracy, GWR was applied using both Gaussian and bisquare weighting function. By using Gaussian and bisquare function, \( RSS \) decrease to \( 7 \times 10^{10} \) and \( 8.43 \times 10^{10} \) respectively. It seems satisfactory since GWR could reduce residual sum of squares of OLR approximately up to 50%. ANOVA was also performed to examine that this model improvement is significant. Results of ANOVA were listed in Table 5. According to this table, ANOVA result for GWR models using both Gaussian and bisquare weighting function were significant, indicated by \( p \)-values. This means GWR performed better on describing variability of per-capita expenditure than OLR did.

<table>
<thead>
<tr>
<th>Weighting function</th>
<th>( df_1 )</th>
<th>( df_2 )</th>
<th>F-statistic</th>
<th>( p )-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gaussian</td>
<td>9.70</td>
<td>21.30</td>
<td>3.04</td>
<td>0.020</td>
</tr>
<tr>
<td>Bisquare</td>
<td>7.35</td>
<td>23.65</td>
<td>3.15</td>
<td>0.045</td>
</tr>
</tbody>
</table>

In addition, to investigate goodness-of-fits of model, Rahmawati (2010) also calculated Pearson product-moment correlation between observed values of per-capita expenditure to correspond fitted values of OLR and GWR. Table 7 summarized the results.
A brief review of geoinformatics analysis on poverty data in Indonesia

### Table 6:
**Pearson product-moment correlation between observed values of per-capita expenditure to it fitted values**

<table>
<thead>
<tr>
<th>Fitted value</th>
<th>( r )</th>
<th>( p )-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>OLR model</td>
<td>0.80</td>
<td>0.00</td>
</tr>
<tr>
<td>GWR model with Gaussian function</td>
<td>0.92</td>
<td>0.00</td>
</tr>
<tr>
<td>GWR model with bisquare function</td>
<td>0.91</td>
<td>0.00</td>
</tr>
</tbody>
</table>

GWR allows model for one region varies to others. Variability of parameter estimates of local model were summarized in Table 7. It is obvious that parameter estimates may be ranged from negative to positive. It means a variable could either decrease or increase per-capita expenditure depends on relative position across villages.

According to two illustrations of GWR application above, it is clear in spatial case – but not always – GWR had better performance than OLR. However, if simplicity is a priority, GWR is not to interesting due to model complexity. For governmental policy, the GWR is more preferable to optimally and appropriately allocate limited resources. Hence, regional priorities to be handled can be recommended by GWR analysis. While the OLR one is usually too broad as it does not facilitate the parameter estimates in specific locations.

### Table 7:
**Summary statistics of parameter estimates of GWR model of per-capita expenditure**

<table>
<thead>
<tr>
<th>Variable</th>
<th>Gaussian weighting function</th>
<th>Bisquare weighting function</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Min</td>
<td>Q1</td>
</tr>
<tr>
<td>Intercept</td>
<td>-171400</td>
<td>14060</td>
</tr>
<tr>
<td>DIST</td>
<td>-8046</td>
<td>-4768</td>
</tr>
<tr>
<td>HFAC</td>
<td>-30150</td>
<td>88950</td>
</tr>
<tr>
<td>ASKES</td>
<td>-3033</td>
<td>-1927</td>
</tr>
<tr>
<td>Intercept</td>
<td>-135000</td>
<td>11250</td>
</tr>
<tr>
<td>DIST</td>
<td>-7322</td>
<td>-4346</td>
</tr>
<tr>
<td>HFAC</td>
<td>-32640</td>
<td>81170</td>
</tr>
<tr>
<td>ASKES</td>
<td>-3469</td>
<td>-1920</td>
</tr>
</tbody>
</table>

(Source: Rahmawati, 2010, pp. 21-22)

### 2.3 Autoregressive Models
The research team also implemented some of the advanced spatial analysis on data dependent on regional factors, i.e. autoregressive models. In this section, a general review was only addressed to Spatial Autoregressive Models (SAR), Spatial Error Models (SEM), and General Spatial Models (SGM). The difference of the three models presented in Table 8. Estimation of parameters in GSM and SAR models were obtained by maximum likelihood approach involving of the areas (Anselin 1988). While the SAR, SEM and SGM are based on the effects of spatial lag (\( \rho \)) and spatial error (\( \lambda \)) among areas. Properties and differences of the three spatial methods mentioned above were presented on Table 8.
These approaches are appealing in the case of spatial autocorrelation which is problematic for ordinary linear regression (OLR), assuming independently distributed errors (Lichstein et al 2002). For spatial data, the SAR model performs better than the OLR, that is when the null hypothesis of no spatial autocorrelation ($\rho=0$) is rejected. In addition, the SAR model performed better than the SEM similar to the results of Mairesse and Mulkay (2008).

<table>
<thead>
<tr>
<th>Model</th>
<th>Parameter estimator</th>
<th>$H_0$</th>
</tr>
</thead>
<tbody>
<tr>
<td>SAR</td>
<td>$\hat{\beta} = (X'X)^{-1}X'y - (X'X)^{-1}\hat{\rho}Wy$</td>
<td>$\rho = 0$</td>
</tr>
<tr>
<td></td>
<td>$\hat{\rho} = (y'Wy)^{-1}y'Wy$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\sigma^2 = \frac{(I-\hat{\lambda}W)(y-X\hat{\beta})(I-\hat{\lambda}W)(y-X\hat{\beta})}{n}$</td>
<td></td>
</tr>
<tr>
<td>SEM</td>
<td>$\hat{\beta} = [X-\hat{\lambda}WX]'(X-\hat{\lambda}WX)^{-1}(X-\hat{\lambda}WX)'(y-\hat{\lambda}Wy)$</td>
<td>$\lambda = 0$</td>
</tr>
<tr>
<td>SGM</td>
<td>$\beta = X'{\Omega}X - 1X'{\Omega}I - \lambda Wy$</td>
<td>$\rho = 0$ or $\lambda = 0$</td>
</tr>
<tr>
<td></td>
<td>$\Omega = (I-\rho W)'(I-\rho W)$</td>
<td></td>
</tr>
</tbody>
</table>


The group compared OLR and autoregressive models on modelling poverty in Indonesia. Response variable in this study was Head Count Index (HCI) of poverty at the district level. The HCI was the percentage of population below poverty line. The explanatory variables were related to education, quality of drinking water, type of work (agriculture and non-agriculture), literacy and type of house (healthy and unhealthy house). General result were presented in Table 9.

<table>
<thead>
<tr>
<th>Covariate</th>
<th>Correlation (p &lt; 0.05)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Education</td>
<td>Negative (tend to reduce HCI)</td>
</tr>
<tr>
<td>Quality of drinking water</td>
<td>Negative (tend to reduce HCI)</td>
</tr>
<tr>
<td>Agricultural work</td>
<td>Positive (tend to increase HCI)</td>
</tr>
<tr>
<td>Non-agricultural work</td>
<td>Negative (tend to reduce HCI)</td>
</tr>
<tr>
<td>Literacy</td>
<td>Non-significant</td>
</tr>
<tr>
<td>Type of house</td>
<td>Negative (tend to reduce HCI)</td>
</tr>
</tbody>
</table>
The data was fitted using SAR, SEM and SGM followed equation in Table 8, the null-hypotheses were all rejected at $\alpha=5\%$, indicating that all models were statistically significant and satisfactory. Table 9 showed that the poverty indicated by the HCI was interlinked with covariates. In addition, spatial models performed better the ordinary regression. More detail explanation on the result were available in the master thesis of the team (Meilisa 2010, Yulianto 2011)

### 3. CONCLUDING REMARK

Poverty in Indonesia can be categorized as regionally distributed data. Therefore, spatial statistics is the mode of choice to analyze this kind of data. Geographically Weighted Regression (GWR) performed better than the ordinary regression based on statistical criteria. In addition, the model provides information at each regional model using the local specific models.

The other approaches for regional data are to implement autoregressive models, i.e. Spatial Autoregressive Models (SAR), Spatial Error Models (SEM), and General Spatial Models (SGM) models. These models provide also better conclusion where the error correlation among region are significant. Many other advanced approaches need to be explored in analysing spatial data including Spatial Generalized Linear Mixed Models and Spatial Small Area Estimation. These complex approaches have been being in the progress of research of the group.

### 4. REFERENCE


RELATIONSHIP AMONG GDP, PER CAPITA GDP, LITERACY RATE AND UNEMPLOYMENT RATE

M. Shafiqur Rahman
Department of Operations Management and Business Statistics
College of Commerce and Economics, Sultan Qaboos University,
Muscat, Sultanate of Oman
Email: srahman@squ.edu.om

ABSTRACT

Gross domestic product (GDP) and per capita GDP (PGDP) are basic measures of the economic performance of a country. The literacy rate is the percentage of people with the ability to read and write. Unemployment rate is the percentage of the total labour force that is unemployed but actively seeking employment and willing to work. This paper investigates the relationship among GDP, PGDP, literacy rate and unemployment rate. It is observed that GDP is not significantly related with PGDP, literacy rate or unemployment rate. There exist significant positive relationship between PGDP and literacy rate but significant negative relationship between PGDP and unemployment rate and between literacy rate and unemployment rate. Therefore taking proper initiative to increase literacy rate of a country will reduce its unemployment rate and increase PGDP resulting development of the country.

1. INTRODUCTION

GDP is a basic measure of the economic performance of a country and is the total market values of all final goods and services produced in a country in a given year. GDP is widely used by economists to measure the health of an economy, as its variations are relatively quickly identified (Samuelson and Nordhaus, 1989). PGDP is the GDP divided by the population as of 1st July of the same year. Differences in PGDP across countries in the world are vast (Knowles and Weatherston, 2007). Literacy has traditionally been described as the ability to read and write. It is a concept claimed and defined by a range of different theoretical fields. The United Nations Educational, Scientific and Cultural Organization (UNESCO) define literacy as the "ability to identify, understand, interpret, create, communicate, compute and use printed and written materials associated with varying contexts. Literacy involves a continuum of learning in enabling individuals to achieve their goals, to develop their knowledge and potential, and to participate fully in their community and wider society." Many policy analysts consider literacy rates as a crucial measure to enhance a region's human capital. This claim is made on the grounds that literate people can be trained less expensively than illiterate people, generally have a higher socio-economic status and enjoy better health and employment prospects. Policy makers also argue that literacy increases job opportunities and access to higher education. In Ireland in 2009, the National Adult Literacy Agency (NALA) commissioned an economist to do a cost benefit analysis of adult literacy training in Ireland. He reported that there were economic gains for the individuals, the companies they worked for, the
Exchequer, as well as the economy, for example, increased GDP, and society at large. The unemployment rate is defined as the number of unemployed persons divided by the labor force, where the labor force is the number of unemployed persons plus the number of employed persons. GDP, per capita GDP, literacy rate and unemployment rate varies from country to country and over time. This paper analyses the GDP, per capita GDP data, literacy rate and unemployment rate of Arab league countries, and OECD countries to test whether there exist any relationship between these variables. This paper also finds out the strategy that a country should take for her development.

2. DATA

The following World Bank data given in Table 1 is the GDP, PGDP, Literacy rate (LR) and Unemployment rate (UR) of Arab league countries obtained from internet.

<table>
<thead>
<tr>
<th>Country</th>
<th>GDP (millions of USD) 2009</th>
<th>GDP per capita (USD) 2009</th>
<th>LR</th>
<th>UR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Algeria</td>
<td>140,848</td>
<td>4,026</td>
<td>72.6</td>
<td>30.0</td>
</tr>
<tr>
<td>Bahrain</td>
<td>20,214</td>
<td>19,455</td>
<td>90.8</td>
<td>2.0</td>
</tr>
<tr>
<td>Comoros</td>
<td>532</td>
<td>798</td>
<td>73.6</td>
<td>14.3</td>
</tr>
<tr>
<td>Djibouti</td>
<td>1,049</td>
<td>1,304</td>
<td>67.9</td>
<td>60.0</td>
</tr>
<tr>
<td>Egypt</td>
<td>187,954</td>
<td>2,450</td>
<td>66.4</td>
<td>11.0</td>
</tr>
<tr>
<td>Iraq</td>
<td>65,838</td>
<td>2,107</td>
<td>77.6</td>
<td>30.0</td>
</tr>
<tr>
<td>Jordan</td>
<td>22,929</td>
<td>3,828</td>
<td>92.2</td>
<td>13.0</td>
</tr>
<tr>
<td>KSA</td>
<td>369,671</td>
<td>14,486</td>
<td>85.5</td>
<td>7.0</td>
</tr>
<tr>
<td>Kuwait</td>
<td>111,309</td>
<td>31,482</td>
<td>94.5</td>
<td>3.0</td>
</tr>
<tr>
<td>Lebanon</td>
<td>33,585</td>
<td>8,706</td>
<td>89.6</td>
<td>11.0</td>
</tr>
<tr>
<td>Libya</td>
<td>60,351</td>
<td>9,529</td>
<td>88.4</td>
<td>30.0</td>
</tr>
<tr>
<td>Mauritania</td>
<td>3,029</td>
<td>975</td>
<td>56.8</td>
<td>32.0</td>
</tr>
<tr>
<td>Morocco</td>
<td>90,815</td>
<td>2,864</td>
<td>56.4</td>
<td>14.0</td>
</tr>
<tr>
<td>Oman</td>
<td>53,395</td>
<td>18,013</td>
<td>86.7</td>
<td>7.0</td>
</tr>
<tr>
<td>Qatar</td>
<td>83,910</td>
<td>68,871</td>
<td>93.1</td>
<td>3.0</td>
</tr>
<tr>
<td>Somalia</td>
<td>2,763</td>
<td>795</td>
<td>51.6</td>
<td>66.0</td>
</tr>
<tr>
<td>Sudan</td>
<td>54,677</td>
<td>1,397</td>
<td>69.3</td>
<td>17.0</td>
</tr>
<tr>
<td>Syria</td>
<td>52,524</td>
<td>2,578</td>
<td>83.6</td>
<td>11.0</td>
</tr>
<tr>
<td>Tunisia</td>
<td>40,168</td>
<td>3,851</td>
<td>78.0</td>
<td>14.0</td>
</tr>
<tr>
<td>UAE</td>
<td>229,971</td>
<td>46,856</td>
<td>90.0</td>
<td>2.0</td>
</tr>
<tr>
<td>Yemen</td>
<td>25,131</td>
<td>1,060</td>
<td>60.9</td>
<td>17.0</td>
</tr>
</tbody>
</table>

The following World Bank data given in Table 2 is the GDP, PGDP, LR and UR of OECD countries obtained from internet.
Table 2:
GDP, PGDP, LR and UR for OECD countries

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Australia</td>
<td>847234.2</td>
<td>39148.0</td>
<td>99.0</td>
<td>4.2</td>
</tr>
<tr>
<td>Austria</td>
<td>332202.7</td>
<td>39849.0</td>
<td>99.0</td>
<td>3.8</td>
</tr>
<tr>
<td>Belgium</td>
<td>394900.0</td>
<td>36879.0</td>
<td>99.0</td>
<td>7.0</td>
</tr>
<tr>
<td>Canada</td>
<td>1295869.2</td>
<td>38883.0</td>
<td>99.0</td>
<td>6.1</td>
</tr>
<tr>
<td>Chile</td>
<td>244217.4</td>
<td>14568.0</td>
<td>96.5</td>
<td>7.8</td>
</tr>
<tr>
<td>Czech Republic</td>
<td>269555.4</td>
<td>25845.0</td>
<td>99.0</td>
<td>4.4</td>
</tr>
<tr>
<td>Denmark</td>
<td>216901.7</td>
<td>39494.0</td>
<td>99.0</td>
<td>3.3</td>
</tr>
<tr>
<td>Estonia</td>
<td>29234.6</td>
<td>21802.0</td>
<td>99.8</td>
<td>5.6</td>
</tr>
<tr>
<td>Finland</td>
<td>200821.2</td>
<td>37795.0</td>
<td>99.0</td>
<td>6.4</td>
</tr>
<tr>
<td>France</td>
<td>2178484.5</td>
<td>33963.0</td>
<td>99.0</td>
<td>7.8</td>
</tr>
<tr>
<td>Germany</td>
<td>3052457.3</td>
<td>37171.0</td>
<td>99.0</td>
<td>7.6</td>
</tr>
<tr>
<td>Greece</td>
<td>337975.4</td>
<td>30077.0</td>
<td>97.1</td>
<td>7.7</td>
</tr>
<tr>
<td>Hungary</td>
<td>207789.5</td>
<td>20700.0</td>
<td>99.4</td>
<td>7.8</td>
</tr>
<tr>
<td>Iceland</td>
<td>12508.0</td>
<td>39166.0</td>
<td>99.0</td>
<td>3.0</td>
</tr>
<tr>
<td>Ireland</td>
<td>189463.7</td>
<td>42644.0</td>
<td>99.0</td>
<td>6.3</td>
</tr>
<tr>
<td>Israel</td>
<td>202302.3</td>
<td>27679.0</td>
<td>97.1</td>
<td>6.1</td>
</tr>
<tr>
<td>Italy</td>
<td>1990547.9</td>
<td>33269.0</td>
<td>98.9</td>
<td>6.8</td>
</tr>
<tr>
<td>Japan</td>
<td>4316608.2</td>
<td>33805.0</td>
<td>99.0</td>
<td>4.0</td>
</tr>
<tr>
<td>Korea</td>
<td>1306387.2</td>
<td>26877.0</td>
<td>99.0</td>
<td>3.2</td>
</tr>
<tr>
<td>Luxembourg</td>
<td>43803.0</td>
<td>89742.0</td>
<td>99.0</td>
<td>4.9</td>
</tr>
<tr>
<td>Mexico</td>
<td>1629595.0</td>
<td>15291.0</td>
<td>92.8</td>
<td>4.0</td>
</tr>
<tr>
<td>Netherlands</td>
<td>705069.5</td>
<td>42887.0</td>
<td>99.0</td>
<td>3.1</td>
</tr>
<tr>
<td>New Zealand</td>
<td>124478.4</td>
<td>29077.0</td>
<td>99.0</td>
<td>4.2</td>
</tr>
<tr>
<td>Norway</td>
<td>288428.4</td>
<td>60480.0</td>
<td>99.0</td>
<td>2.5</td>
</tr>
<tr>
<td>Poland</td>
<td>688458.2</td>
<td>18062.0</td>
<td>99.3</td>
<td>7.2</td>
</tr>
<tr>
<td>Portugal</td>
<td>265100.0</td>
<td>24957.0</td>
<td>94.9</td>
<td>8.5</td>
</tr>
<tr>
<td>Slovakia Republic</td>
<td>125638.1</td>
<td>23245.0</td>
<td>99.0</td>
<td>9.5</td>
</tr>
<tr>
<td>Slovenia</td>
<td>59083.9</td>
<td>29221.0</td>
<td>99.7</td>
<td>4.4</td>
</tr>
<tr>
<td>Spain</td>
<td>1512485.1</td>
<td>33173.0</td>
<td>97.9</td>
<td>11.4</td>
</tr>
<tr>
<td>Sweden</td>
<td>363958.0</td>
<td>39475.0</td>
<td>99.0</td>
<td>6.2</td>
</tr>
<tr>
<td>Switzerland</td>
<td>351514.8</td>
<td>45586.0</td>
<td>99.0</td>
<td>3.2</td>
</tr>
<tr>
<td>Turkey</td>
<td>1063519.1</td>
<td>14962.0</td>
<td>88.7</td>
<td>9.7</td>
</tr>
<tr>
<td>UK</td>
<td>2260520.4</td>
<td>36817.0</td>
<td>99.0</td>
<td>5.7</td>
</tr>
<tr>
<td>USA</td>
<td>14296900.0</td>
<td>46901.0</td>
<td>99.0</td>
<td>5.8</td>
</tr>
</tbody>
</table>

3. METHODOLOGY

In order to study the relationship among GDP, PGDP, literacy rate and unemployment rate, we first calculate the Pearson’s Product moment correlation coefficient (Johnson and Bhattacharyya, 2001) pair-wise for all four variables using the following formulae and presented in Table 3 and Table 5 for Arab League countries and OECD countries respectively.
In order to test the significance of the correlation coefficient the following null (H$_0$) and alternative (H$_1$) hypotheses are considered.

\[ H_0: \rho = 0 \]
\[ H_1: \rho \neq 0 \]

where \( \rho \) is the population correlation co-efficient between two variables.

The appropriate test statistic to test the above hypothesis is

\[ t = \frac{r \sqrt{n-2}}{\sqrt{1-r^2}} \]

The values of the test statistic and P-values are given in Table 4 and Table 6. In order to study the relationship among GDP, PGDP, literacy rate and unemployment rate, we also calculate the Spearman’s rank correlation (Conover 1980) pair-wise for all four variables using the following formulae and presented in Table 8 for Arab League countries.

\[ \rho_s = 1 - \frac{6T}{n(n^2-1)} \]

where, \( T = \sum_{i=1}^{n} [R(X_i) - R(Y_i)]^2 \), \( R(X_i) \) be the rank of \( X_i \) and \( R(Y_i) \) be the rank of \( Y_i \) where \( i = 1, 2, 3, \ldots, n \).

4. RESULTS AND DISCUSSION

In order to study the relationship among GDP, PGDP, Literacy rate (LR) and Unemployment rate (UR) first we construct pair wise scatter diagram for Arab league countries and presented in the following Graphs (Graph 1 to graph 6).
Graph 2: GDP VS LR for Arab league countries

Graph 3: GDP VS UR for Arab league countries

Graph 4: PGDP VS LR for Arab league countries
Relationship among GDP, Per Capita GDP, Literacy Rate and Unemployment Rate

From the Graphs it seems there is no relation or weak relation between PGDP and GDP, GDP and LR, GDP and UR. There may be positive relationship between PGDP and LR, negative relationship between PGDP and UR and between LR and UR. Similar pattern also observed for OECD countries. Then we find the correlation matrix for GDP, PGDP, LR and UR for Arab league countries and presented in Table 3.

Table 3:
Correlation Matrix for GDP, PGDP, LR and UR for Arab league countries

<table>
<thead>
<tr>
<th></th>
<th>GDP</th>
<th>PGDP</th>
<th>LR</th>
<th>UR</th>
</tr>
</thead>
<tbody>
<tr>
<td>GDP</td>
<td>1</td>
<td>0.3217</td>
<td>0.2445</td>
<td>-0.3823</td>
</tr>
<tr>
<td>PGDP</td>
<td>0.3217</td>
<td>1</td>
<td>0.5961</td>
<td>-0.4815</td>
</tr>
<tr>
<td>LR</td>
<td>0.2445</td>
<td>0.5961</td>
<td>1</td>
<td>-0.6236</td>
</tr>
<tr>
<td>UR</td>
<td>-0.3823</td>
<td>-0.4815</td>
<td>-0.6236</td>
<td>1</td>
</tr>
</tbody>
</table>

Table 4 represents the test statistic values and P-values for testing the significance of correlation coefficient between four variables of Arab league countries.
There exist significant positive relationship between PGDP and LR, significant negative relationship between PGDP and UR and between LR and UR. However, no significant relationship exists between PGDP and GDP, GDP and LR, GDP and UR.

Table 5 represents the correlation matrix for GDP, PGDP, LR and UR of OECD countries.

### Table 5: Correlation Matrix for GDP, PGDP, LR and UR for OECD countries

<table>
<thead>
<tr>
<th></th>
<th>GDP</th>
<th>PGDP</th>
<th>LR</th>
<th>UR</th>
</tr>
</thead>
<tbody>
<tr>
<td>GDP</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PGDP</td>
<td>0.1130</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LR</td>
<td>0.0345</td>
<td>0.3985</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>UR</td>
<td>0.0309</td>
<td>-0.3750</td>
<td>-0.3418</td>
<td>1</td>
</tr>
</tbody>
</table>

Table 6 represents the test statistic values and P-values for testing the significance of correlation coefficient between four variables of OECD countries.

### Table 6: Test statistic and P-values for testing the significance of correlation coefficient

<table>
<thead>
<tr>
<th>Relation between</th>
<th>r</th>
<th>t-ratio</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>PGDP and LR</td>
<td>0.5961</td>
<td>3.2361</td>
<td>0.0043</td>
</tr>
<tr>
<td>PGDP and GDP</td>
<td>0.3217</td>
<td>1.4809</td>
<td>0.1550</td>
</tr>
<tr>
<td>LR and GDP</td>
<td>0.2445</td>
<td>1.0992</td>
<td>0.2854</td>
</tr>
<tr>
<td>PGDP and UR</td>
<td>-0.4815</td>
<td>-2.3946</td>
<td>0.0271</td>
</tr>
<tr>
<td>GDP and UR</td>
<td>-0.3823</td>
<td>-1.8033</td>
<td>0.0872</td>
</tr>
<tr>
<td>LR and UR</td>
<td>-0.6236</td>
<td>-3.4769</td>
<td>0.0025</td>
</tr>
</tbody>
</table>

There exist significant positive relationship between PGDP and LR, significant negative relationship between PGDP and UR and between LR and UR. However, no significant relationship exists between PGDP and GDP, GDP and LR, GDP and UR.

We also apply non-parametric test for more confirmation. The values of the four variables GDP, PGDP, LR and UR for Arab league countries are replaced by their corresponding ranks and presented in Table 7.
Spearman’s rank correlation between GDP and LR, PGDP and LR, PGDP and UR, LR and UR are calculated for Arab league countries and tested. The test results are given in Table 8.

### Table 7:
Ranks of GDP, PGDP, LR and UR for Arab league countries

<table>
<thead>
<tr>
<th>GDP</th>
<th>PGDP</th>
<th>LR</th>
<th>UR</th>
</tr>
</thead>
<tbody>
<tr>
<td>18</td>
<td>13</td>
<td>8</td>
<td>17</td>
</tr>
<tr>
<td>5</td>
<td>18</td>
<td>18</td>
<td>1.5</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>9</td>
<td>13</td>
</tr>
<tr>
<td>2</td>
<td>5</td>
<td>6</td>
<td>20</td>
</tr>
<tr>
<td>19</td>
<td>8</td>
<td>5</td>
<td>8</td>
</tr>
<tr>
<td>14</td>
<td>7</td>
<td>10</td>
<td>17</td>
</tr>
<tr>
<td>6</td>
<td>11</td>
<td>19</td>
<td>10</td>
</tr>
<tr>
<td>21</td>
<td>16</td>
<td>13</td>
<td>5.5</td>
</tr>
<tr>
<td>17</td>
<td>19</td>
<td>21</td>
<td>3.5</td>
</tr>
<tr>
<td>8</td>
<td>14</td>
<td>16</td>
<td>8</td>
</tr>
<tr>
<td>13</td>
<td>15</td>
<td>15</td>
<td>17</td>
</tr>
<tr>
<td>4</td>
<td>3</td>
<td>3</td>
<td>19</td>
</tr>
<tr>
<td>16</td>
<td>10</td>
<td>2</td>
<td>11.5</td>
</tr>
<tr>
<td>11</td>
<td>17</td>
<td>14</td>
<td>5.5</td>
</tr>
<tr>
<td>15</td>
<td>21</td>
<td>20</td>
<td>3.5</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>1</td>
<td>21</td>
</tr>
<tr>
<td>12</td>
<td>6</td>
<td>7</td>
<td>14.5</td>
</tr>
<tr>
<td>10</td>
<td>9</td>
<td>12</td>
<td>8</td>
</tr>
<tr>
<td>9</td>
<td>12</td>
<td>11</td>
<td>11.5</td>
</tr>
<tr>
<td>20</td>
<td>20</td>
<td>17</td>
<td>1.5</td>
</tr>
<tr>
<td>7</td>
<td>4</td>
<td>4</td>
<td>14.5</td>
</tr>
</tbody>
</table>

This test also indicates that there exist significant positive relation between PGDP and LR, significant negative relation between PGDP and UR and between LR and UR.

### Table 8:
Spearman’s rank correlation test results for Arab league countries

<table>
<thead>
<tr>
<th>Relation between</th>
<th>$r_s$</th>
<th>5% Critical value</th>
<th>Conclusion</th>
</tr>
</thead>
<tbody>
<tr>
<td>GDP and LR</td>
<td>0.2442</td>
<td>0.4351</td>
<td>Insignificant</td>
</tr>
<tr>
<td>PGDP and LR</td>
<td>0.8260</td>
<td>0.4351</td>
<td>Significant</td>
</tr>
<tr>
<td>PGDP and UR</td>
<td>-0.7847</td>
<td>-0.4351</td>
<td>Significant</td>
</tr>
<tr>
<td>LR and UR</td>
<td>-0.7198</td>
<td>-0.4351</td>
<td>Significant</td>
</tr>
</tbody>
</table>

5. COMMENTS AND CONCLUSION

There exist significant positive relationship between PGDP and LR but significant negative relationship between PGDP and UR and between LR and UR. GDP is not significantly related with PGDP, LR or, UR. Therefore, PGDP plays a significant role for changing LR and UR. In order to develop a country the government of that country can
easily take some proper initiative for example increase education budget, making primary/secondary education compulsory, etc. to increase LR. Increasing LR will reduce UR and increase PGDP resulting development of the country.

REFERENCES

THE CONSUMPTION PATTERN OF HOUSEHOLD FOOD-AWAY-FROM-HOME: PAKISTAN

Ammara Nawaz Cheema¹ and Naheeda Akhtar²

¹ Department of Mathematics, Air University, Islamabad, Pakistan.
Email: ammara_qau@yahoo.com
² Department of Statistics, University of Wah, Wah Cantt, Pakistan.
Email: nadih87@yahoo.com

ABSTRACT

In the modern era, people are usually spending a busy life, their priorities are changed and they are utilizing the modern facilities rapidly. This study characterizes the household food-away-from-home (FAFH) expenditure pattern in selected areas’ of Pakistan (Rawalpindi and Wah). Specifically, a Logit model was estimated to quantify the responsiveness of the Surveyed household FAFH expenditures to changes in their income and selected household demographic characteristics.

The findings of survey consisting of 500 observations, demonstrate that, Households located in urban areas, with more family members, and whose household head is young and male had generally higher levels of FAFH expenditure.

KEYWORDS:
Food away from home (FAFH), Logit model, Expenditure; consumption.

1. INTRODUCTION

One of the largest recent changes in consumer food purchasing behavior is the trend towards greater consumption of food eaten outside the home. Usually people in Pakistan cooked meals at home with grain, raw vegetables, and meat produced at home, or purchased from food stores. Along with rapid income growth and the modernized changes, the attitude of Pakistani consumers has also changed just like the consumers of the development countries. Most of Pakistani females are working; spend most of the time outside from home. So they prefer to purchase meals from restaurants, cafeterias etc or eating meals in restaurants, cafeterias, and dining halls.

This study focuses on whether demographic and socioeconomic factors have detectable effects on FAFH expenditures in the areas’ of Pakistan (Rawalpindi and Wah). This question is of interest because previous studies suggest that increasing household income, education, female labor participation and changing lifestyle especially in developing countries increased household FAFH consumption share but decreased food at home consumption share. However, no study to our knowledge has examined the effect of income, education, job status and household size on Pakistan FAFH consumption. The data for this research were obtained from personal interviews of sample households of the Rawalpindi and Wah. The findings of this study generally...
indicate that restaurant facilities, employment of wife and education, composition of household, and income are statistically significant determinants of FAFH consumption.

The food away from home expenditures implies the creation of demand for high value and specialty food products and restaurant services. A better understanding of the factors associated with FAFH expenditures has become increasingly important to understanding changes in the food commodity market, forecasting food demand, anticipating the implications of changes in eating patterns on diet and food imports, and the design of effective marketing programs for both domestic and international restaurant business.

The paper is developed in the following way. In the first part, we present the review of the literature, in the second part; we explain our methodology and description of data. Then, we present the estimated model and results. Finally, the conclusions are given.

2. REVIEW OF THE LITERATURE

Previous studies have investigated household food expenditures at home in different countries e.g. Heng and Guan (2007) examine household expenditure patterns on food-away-from-home (FAFH) using the Tobit model on data from the Malaysian Household Expenditure Survey (1998/99). It is indicated by the results that the Chinese population, the urban residents, or those with higher monthly household Income have significantly higher FAFH expenditures than their non-Chinese, rural, or lower household income cohorts, ceteris paribus. In addition, other socio-demographic characteristics such as age, gender, household size, or even education do not affect total monthly household expenditures on FAFH in a statistically significant manner.


McCracken and Brandt (1987) used the quarterly data of spring 1977-78 from the nationwide Food Consumption Survey (U.S. Department of Agriculture), to study the determinants of total household expenditures on food away from home by type of facilities. They used the Tobit model with the variables, household income, time value, size of the household and found that these factors varied importantly between conventional restaurants, fast-food facilities, and other commercial establishments.

Yen (1993) studies the patterns of working wives Food Away From Home expenditures with a Box-Cox double hurdle model. Byrne, Capp, and Saha (1996) use the generalized Heckman two-step estimation procedure, which is more flexible than a Tobit model, to study the U.S. FAFH expenditure pattern.

Elsner (1999) analyzed food expenditure pattern of Russia by using a two-stage budgeting system. He estimated the total expenditure allocation on food and non-food by using Working's Engel model in the first stage. The estimated income elasticity for demand of food was 0.81 in Russia. Also, the estimated income elasticity for demand of food among Russian households in rural areas and urban areas was to be 0.98 and 0.78 respectively.
Obayelu et al. (2009) examines food consumption patterns between the rural and urban households in the North-central Nigeria using data obtained from households’ seven days memory recall. The results of the descriptive statistics revealed heterogeneity in consumption and expenditure patterns across households in rural and urban areas. Urban residents purchase 37.9% of the food they consume, while families in rural areas purchase only 26.6%. The actual consumption was based on income and their membership of social group.

3. METHODOLOGICAL CONSIDERATION

3.1 Data on households’ food consumption, food expenditure and related variables

After reviewing different research methods which are commonly in use, we found survey research to be more appropriate for our research study. Sampling, designing questions and interview are the basic components of survey approach (see Fowler, 1993). Corbetta (2003), say that survey is the widest spread quantitative research techniques. Here the population refers to the individual being researched, so a major step in social research is to define population clearly that consequently may help in selection of a representative sample for inferring characteristic of the part (sample) that close to the whole (population) (Labovitz and Hagedorn, 1981). The proposed sampling technique of the present study has in built method of estimating sampling errors as standard error. However estimates are blown up through the application of ratio estimation method with the help of two auxiliary variables, age and sex, as controlled variables as to match with the actual population count emerged through quick count listing.

A three-stage stratified systematic random sample was drawn with probability proportionate to block size in term of number of households per block. There are 250 observations related to FAFH expenditure are in urban areas and in rural areas of Wah and 163 observations related to FAFH expenditure are in urban areas and in rural areas of Islamabad Capital Territory with 20 social, economic and demographic characteristics of FAFH are picked up.

3.2 The Logit Model

Consider a relationship between expenditure on food away from home Y, and a matrix of socio-demographic factors and income, X. A linear regression model yields

\[ Y = X\beta + e \]

where “\( e \)” represents unknown error terms and “\( \beta \)” are parameters to be estimated.

Since in consumption data for FAFH, zero observations are present, the linear regression model can’t be employed. In such a case of binary dependent variable a logit model is specified. Here we have also considered the logit model to identify the relationship between socio-demographic variables and the expenditures on food away from home.

The probability of taking food away from home is modeled as

\[ p(y_i > 0) = \frac{e^{\beta X_i}}{1 + e^{\beta X_i}} \]  

(1)
where household variables include the number of individuals in the household, the geographical region the household is located within, whether the household is located in an urban or rural area, and household income. Demographic variables include age, typical number of hours spent working per week, typical number of hours spent watching television per day, sex, job status and the highest education level achieved. Other variables not really household or individual related include the month and day of the week the food item was consumed, the name of the meal (i.e. breakfast, lunch, etc.

4. ESTIMATED MODEL AND RESULTS

The logit model present in equation (1) is estimated using the maximum likelihood method in E-views and results are presented in table (1).

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>Std. Error</th>
<th>z-Statistic</th>
<th>Prob.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>1.407218</td>
<td>0.878884</td>
<td>1.601141</td>
<td>0.1093</td>
</tr>
<tr>
<td>Total Expenditure</td>
<td>3.52E-05</td>
<td>7.85E-06</td>
<td>4.488244</td>
<td>0.0000</td>
</tr>
<tr>
<td>Household Size</td>
<td>0.032353</td>
<td>0.061368</td>
<td>0.527197</td>
<td>0.5981</td>
</tr>
<tr>
<td>Urban-Rural Location</td>
<td>-0.123181</td>
<td>0.252556</td>
<td>-0.487736</td>
<td>0.6257</td>
</tr>
<tr>
<td>Gender</td>
<td>-0.212423</td>
<td>0.378107</td>
<td>-0.561806</td>
<td>0.5742</td>
</tr>
<tr>
<td>Marital Status</td>
<td>0.014730</td>
<td>0.280869</td>
<td>0.052445</td>
<td>0.9582</td>
</tr>
<tr>
<td>Age</td>
<td>-0.030495</td>
<td>0.011583</td>
<td>-2.632716</td>
<td>0.0085</td>
</tr>
</tbody>
</table>

Table 1 reports the standard Logit model estimates for the FAFH expenditure model. With the exception of household size, urban-rural location, Gender of the head of household and marital status of the head of household, the rest of the explanatory variables are significant at the 1% level. The signs are also consistent with expectations. That is, income has a significant positive influence on FAFH expenditure. The size of the household also shows a positive influence on FAFH expenditure, although not significant. That is, households with more family members spent more on FAFH compared to households with fewer members.

Moreover, rural location of households has a negative influence on FAFH expenditure. This household location effect can be explained from both the demand and supply sides. That is, urban location is highly associated with higher income, which has a positive influence on FAFH expenditure on the demand side. On the supply side, there are more institutional food suppliers in urban locations, so this sector is larger in urban compared to rural areas.

When households are headed by males who are older, this shows a negative influence on FAFH expenditure. The marital status of the head of household is shows a positive influence on FAFH expenditure, although not significant.

Table 2 reports the standard Logit model estimates for the FAFH expenditure model. With the exception of household size, urban-rural location, Gender of the head of household, age and marital status of the head of household, the rest of the explanatory variables are significant at the 1% level. The signs are also consistent with expectations. That is, income has a significant positive influence on FAFH expenditure. The size of the
household also shows a positive influence on FAFH expenditure, although not significant. That is, households with more family members spent more on FAFH compared to households with fewer members.

Table 2: Logit model estimates of FAFH expenditure model for Wah

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>Std. Error</th>
<th>z-Statistic</th>
<th>Prob.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>-0.511530</td>
<td>1.179475</td>
<td>-0.433693</td>
<td>0.6645</td>
</tr>
<tr>
<td>Total Expenditure</td>
<td>9.35E-05</td>
<td>1.90E-05</td>
<td>4.924139</td>
<td>0.0000</td>
</tr>
<tr>
<td>Household Size</td>
<td>-0.082782</td>
<td>0.079765</td>
<td>-1.037827</td>
<td>0.2994</td>
</tr>
<tr>
<td>Urban-Rural Location</td>
<td>0.265445</td>
<td>0.317136</td>
<td>0.837007</td>
<td>0.4026</td>
</tr>
</tbody>
</table>

Moreover, urban location of households has a positive influence on FAFH expenditure. This household location effect can be explained from both the demand and supply sides. That is, urban location is highly associated with higher income, which has a positive influence on FAFH expenditure on the demand side. On the supply side, there are more institutional food suppliers in urban locations, so this sector is larger in urban compared to rural areas.

When households are headed by males who are younger, this shows a negative influence on FAFH expenditure. The marital status of the head of household is shows a positive influence on FAFH expenditure, although not significant.

Table 3: Logit model estimates of FAFH expenditure model for Rawalpindi

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>Std. Error</th>
<th>z-Statistic</th>
<th>Prob.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>3.591173</td>
<td>1.639861</td>
<td>2.189925</td>
<td>0.0285</td>
</tr>
<tr>
<td>Total Expenditure</td>
<td>2.58E-05</td>
<td>9.67E-06</td>
<td>2.668997</td>
<td>0.0076</td>
</tr>
<tr>
<td>Household Size</td>
<td>0.163520</td>
<td>0.129436</td>
<td>1.263325</td>
<td>0.2065</td>
</tr>
<tr>
<td>Urban-Rural Location</td>
<td>-1.989076</td>
<td>0.632102</td>
<td>-3.146763</td>
<td>0.0017</td>
</tr>
<tr>
<td>Head of Household</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gender</td>
<td>-0.558960</td>
<td>0.623716</td>
<td>-0.896177</td>
<td>0.3702</td>
</tr>
<tr>
<td>Marital Status</td>
<td>-0.149299</td>
<td>0.458762</td>
<td>-0.325440</td>
<td>0.7448</td>
</tr>
<tr>
<td>Age</td>
<td>-0.031820</td>
<td>0.019810</td>
<td>-1.606218</td>
<td>0.1082</td>
</tr>
</tbody>
</table>

Table 3 reports the standard Logit model estimates for the FAFH expenditure model. With the exception of household size, Gender of the head of household, age and marital status of the head of household, the rest of the explanatory variables are significant at the 1% level. The signs are also consistent with expectations. That is, income has a significant positive influence on FAFH expenditure. The size of the household also shows a positive influence on FAFH expenditure, although not significant. That is, households with more family members spent more on FAFH compared to households with fewer members.
Moreover, rural location of households has a negative influence on FAFH expenditure. This household location effect can be explained from both the demand and supply sides. That is, urban location is highly associated with higher income, which has a positive influence on FAFH expenditure on the demand side. On the supply side, there are more institutional food suppliers in urban locations, so this sector is larger in urban compared to rural areas.

When households are headed by males who are older, this shows a negative influence on FAFH expenditure. The marital status of the head of household is shows a positive influence on FAFH expenditure, although not significant.

5. CONCLUSION

This study characterized the household FAFH expenditure pattern in selected areas’ of Pakistan (Rawalpindi and Wah). We found that based on the Dec 2010 survey data, only 250 and 163 of households interviewed reported positive FAFH expenditures in Wah & Rawalpindi respectively. On average, household FAFH expenditure accounted for 5% to 8% of total food expenditure. In general, larger households and those located in urban areas spent more on FAFH.

REFERENCE

REDUNDANT USE OF MOBILE AFFECTS STUDENT’S GPA

Ammara Nawaz Cheema
Department of Mathematics, Air University, Islamabad, Pakistan
Email: ammara_qau@yahoo.com

ABSTRACT

Use of mobile phone during lecture may have its implications at student’s GPA. This paper explains a small but considerable issue, that use of mobile during lecture and routine classes affects students’ grades and GPA. Unanimously students don’t bother about this fact, but somehow it is true and our research determined that it is the one of the critical factor which affects students GPA. This factor is identified and categorized; the relevance of such factor with students GPA is examined in a survey consisting of 150 observations. The findings demonstrate that its effect on students class behavior, attention towards lecture subject, non serious attitude while attending calls and receiving messages, completely disturbs class environment. These are some basic reasons which are affecting students GPA at end of the day. Though, this act is the least important. Many other reasons are more influential of lesser grades but this factor is becoming considerable for university students specially. The study reflects affect of mobile use during lecture on students’ GPA and the finding may be of interest for the researcher and parishioners.

OVERVIEW

In many high schools, the use of cell phones has been banned. In some schools, the mere possession of a cell phone can result in trouble for teens, including losing whole grades, and the teachers obtaining possession of the phones. Cell phones have become a big issue in schools, and controversy has arisen all over the place. Many parents feel that their children need to carry their phones with them for after school rides, allowing communication between their teen and themselves, and for simple safety reasons, as many high-school students are just receiving their drivers licenses and parents feel they are a safety precaution in case of an emergency (Mathews 2006). But the schools take a different perspective on the use of phones in school, since many students have found ways to alert their friends of test answers through the use of text messaging, which can be done discreetly and silently. Since many teachers use the same tests for all class periods, kids with the tests in the first hours of the day are able to record the answers and pass them along, or even charge other students, for the answers. Since cheating has become such a large threat in schools, with the expanding use of the internet and the availability of so many resources, as well as the ease of use through cell phones and passing information, the ban, in some circumstances, does not seem unfair.

Students, who are receiving higher grades through cheating, make it far more difficult for other students who are truly putting their best effort forth. Not only are they potentially getting better grades, but they are also raising the grade standards for the high
Redundant Use of Mobile Affects Student’s GPA

school as well as the possibility of getting into certain colleges who require higher grade point averages. When many students start receiving high grade point averages, it raises the standard for all students to meet. The use of cell phones in school may not directly affect the student’s grades, but in some schools where the ban has been put in place, just the possession of a cell phone seen by a teacher could result in a dock of a full letter grade. Other schools are trying to prevent the threat of cheating through test messaging and other methods by simply not allowing them. But the threat is still there, you’ll always have cheaters, and you’ll always have those who benefit from doing so and breaking the rules that were put in place for a reason.

In our society, it is hard to imagine what life was like before cell phones. One of the ways that cell phones have revolutionized the way we live is cell phones in school. Some people think that phones have no place in schools, but banning cell phones in school would be like going back in time to a less convenient and safe time. It seems a little ridiculous for children in elementary school to have cell phones, but at the end of the day, this is a safety tool for kids and provides peace of mind to parents everywhere. In a world, where unfortunately, we constantly hear about school violence, the cell phone provides an instant connection, not only to parents, but to authorities. If a student were to pull a gun at a school that banned cell phones, help would likely take longer to reach the school. Additionally, a lot of parents panic when something like that happens and cell phones provide instant connections to their children (Montano 2008). This helps things run more smoothly in case of emergencies (Mathews 2006).

Over 4 billion people carry cell phones worldwide according to a report from the United Nations. The use of cell phone has exploded since this practical device first hit the scene in 1983. With this type of growth, there are bound to be effects on people and the society as a whole. Some effects are positive while others can present challenges.

The idea is to get the word out to all the students quickly and effectively in the event of an emergency like that. They could get the school mobilized and away from the buildings or classrooms that are in immediate danger. A lot of people dwell on the small problems that come with children having cell phones in school, like using texting to cheat or children playing games on their phone instead of paying attention to the class. But children don’t need technology to cheat and to be distracted. If a child is going to cheat, they’ll find a way to do it even without cell phones; the way people cheated back in the day when cell phones weren’t around. Similarly, not having cell phones to distract them from class does not mean they will pay attention. They’ll just find another way to be distracted. And as a teacher, if you’re adamant about cell phone use, you can impose no cell phone rules for your classroom. You can usually tell when a student is doing something they’re not supposed to be doing and if that bothers you, you can discipline them. In addition, cell phones provide the ultimate convenience. You can arrange where to pick up your children or give them an easy way to communicate to you that they’re going home with a friend or to the mall after school. It’s a way for parents to easily keep track of their children. All the reasons not to allow cell phone use in school seem petty and are dwarfed by all the reasons why it’s beneficial.

Cell phones can be a huge distraction. Teen-agers and middle-school children seem to feel that if they are awake, they should be texting someone on their cell phones. New
rules have to be applied in families. You must constantly remind your children that cell phone usage is not allowed at the dinner table and during family time. Cell phone usage is almost epidemic in our society. People seem unable to walk down the street or sit outside enjoying the sunshine without speaking or texting on their cell phones. Many people argue that talking on a cell phone while driving is not any more disruptive than carrying on a casual conversation with an occupant of the vehicle. According to the American Automobile Association, the driver may not be distracted by a lighthearted, casual conversation, but he may be paying less attention to his driving if the conversation is intense, such as during a business conversation or an argument with someone.

There are many arguments for banning cell phones, but for the most part they all have to do with the disruption of the learning process. Many kids are text messaging or playing games on their phones during class, instead of paying attention. This obviously leads to less comprehension of classroom material and worse grades. Others are taking pictures that they later post on the Internet. A lot of these pictures are used to embarrass fellow classmates. You might have heard stories about fights that started as a result of something posted online. It happens a lot these days, and cell phones only add to the problem.

Personal responsibility would solve most of these problems. If children choose to play around on a cell phone during class and suffer bad grades, they should be held accountable, as it was their fault they didn't learn. Of course, this kind of accountability will never happen. First of all, they're kids, and kids are rarely mature enough to make smart decisions when it comes to this kind of thing. Secondly, many parents refuse to admit their kids could be doing anything wrong, and will blame the teacher for bad grades, even if a cell phone is to blame.

The main problem that people have with students being able to use their cell phones during school hours is that students can use them to cheat. If it were not for that, most people would not have an issue, but since students are able to do this, it has become a problem (Smithe 2008). The problem is that high schools allow students to have cell phones, whether or not they are allowed to use them during school hours, and most students will use them whether or not they are allowed to.

**METHODOLOGY & RESEARCH INSTRUMENT**

After reviewing different research methods which are commonly in use, we found survey research to be more appropriate for our research study. For survey approach, determining sampling, developing questionnaires and interviews are the basic components (Fowler, 1993) as it is the widest spread quantitative research technique Corbetta (2003). As the population refers to individuals being researched, a major step in social research is to define population clearly that consequently may help in selection of a representative sample for inferring characteristics of the population (Labovitz and Hagedorn, 1981). The proposed sampling technique of the present study has in-built method for estimating sampling errors as standard error. However, estimates can be inflated through the application of ratio estimation method with the help of two auxiliary variables, i.e. age and sex used as controlled variables, as to match with the actual population count emerging from quick count listing.
Nearly 5.5 percent sample was selected. A three-stage stratified systematic random sample with probability proportionate to block size in terms of number of department per block was drawn. There are 150 observations related to mobile users during class lectures of universities of Islamabad Capital Territory, whereas 15 questions of mobile use were picked up for the research purpose.

Our study relates usage of a mobile to its impact on students’ GPA. We design a questionnaire and use five level likert scales. For proper data entries we use excel and for data analysis we use SPSS software. To get the most frequent result we use mode and graphical representations for better understanding.

**DATA ANALYSIS AND DISCUSSION**

**Using of mobile during lecture**

According to our research carried by questionnaire that most of the students use mobile during the lecture and in routine classes.

**Bad impact on studies**

Results stand with our hypotheses; maximum of students are agreed with the bad impact of mobile on studies and only few disagreed.
Useful during lecture
Students disagree with this statement, and stand with our hypotheses.

Lesser the use of mobile phone results higher the students’ GPA
Students do agree with our assumption, that mobile use is not beneficial.

Strictness is main cause of bunker
Students disagree with this statement, because students bunk classes rarely for completing their assignments, preparing and revising quizzes sometimes for their enjoyment.
**Students attention divert by mobile use**
Students agree with us, because when you use mobile it diverts your attention from studies and results a decrease in GPA at end of the semester.

**Mobile effect students’ grade**
Students agree with this assumption, that mobile use effect grades.

**Strictness only during exams and test**
University can solve this issue of using cell phones during lectures by restricting its use during tests and exams only.
Also divert other students’ attention
When a student uses mobile during lecture it not only affects his own concentration but also attention of other students from lecture.

Disturb teacher attentions
Use of mobile during lecture also diverts attention of teacher. Teachers cannot ignore this disturbance. So it affects teachers as well as other students.

Students are mature enough to understand pros and cons of their acts
Students agree that they are self responsible for subject’s side effect. They know well what is good for them.

CONCLUSION
By keeping in mind problem statement and according to the current situation of students in universities we will proof that this statement “Use of mobile phone during lectures has bad impact on students GPA” is true. If students use mobile during lecture he will lose his concentration. Physically he will be present in class but mentally his mind will be busy in using mobile, chatting with friends. And then continuity of this non serious behavior gives unpleasant results at the end. Though, this act is the least important. Many other reasons are more influential of lesser grades but this factor is becoming considerable for university students specially. Students do agree with the point.
They seem to be less interested for the results but they agree with the disadvantages caused by excessive use of mobiles during lectures. It not only affects that student but also fellow students in class. Teacher also gets disturb. So suitable measure should be taken to resolve the issue. Conceptual studies basically depend on concentration and attention towards lecture. When you lose your concentration even for a minute entire concept become foggy and this shortage is un-ignorable.

REFERENCES

FUNCTIONAL TRAINING AND ITS IMPACT ON THE DEVELOPMENT OF STATISTICAL AWARENESS

Waleed Ameen Abd Elkhalek Mohamed
Master of Curriculum and Instruction
Statistician at National Center for Statistical Training
CAPMAS, Salah Salem Street, Nasr City, Cairo, Egypt
Email: woo2005@yahoo.com; kant2011xp@hotmail.com; pres_capmas@capmas.gov.eg

ABSTRACT

Many researchers in education and psychology support the theory that students learn by actively building or constructing their own knowledge and making sense out of this knowledge, also they believe that Individuals can be construct new knowledge internally by transforming, organizing, and reorganizing previous knowledge (Cobb 1994; Greeno, Collins, and Resnick 1996) as well as externally, through environmental and social factors that are influenced by culture, language, and interactions with others that is the core of functional training.

The functional training based on basic premise: that education is life, not preparation for life, therefore trainees feels that they accept to study statistics because it face the problems of life and give them the real indicators.

In brief Functional Training is One of educational approaches that emphasize the link between what trainees learn (knowledge, facts and information) with their daily life to help them applying what they have studied theoretically on their life, to achieve maximum benefit and to make the material meaningful in real life.

OBJECTIVES OF THE STUDY

1. Describes how to use The Functional Training in teaching statistics.
2. Clarify the role of The National Center for Statistical Training in the process of raising statistical awareness through statistical training programs
3. Demonstrate the importance of statistical awareness in various fields such as (policy-making, Scientific Research, Statistics and predication and Strategic Planning).
4. Describes how the field researchers deal with types of respondent (Silent, Jabber, Sceptic, nervous, special needs) with function way.
5. Measure degree of satisfaction of the trainee for the training program (SPSS program) using Likert scale.

Methodology of the study:
- Using descriptive analytical method by studying (books - Literature Internet-related subject of the study (Likert scale and trends), as was the use of a survey form for this purpose.
- Select a sample of trainees who attending courses at (CAPMAS).
Do some statistical treatments using the SPSS program.
Calculate validity and reliability of the survey questions using a factor (alpha Cronbach).
Calculate direction of the trainees in accordance with the methodology for the Likert and using the SPSS.

6. Recommendation.

KEY WORDS
Functional Training; statistical awareness; satisfaction of the trainee; trend.

INTRODUCTION
There is a famous for the Greek philosopher Herakulaitts 475-540 BC. which he says "We do not go down the river twice," the significance of this phrase that existence in continuous change because the development is the law of thought, existence and life in general.

The most important feature in contemporary life is the scientific and technological revolution and the resulting shifts in various aspects of human life in the economy - media - politics - education ... etc.

There is no doubt that education and training plays an important role in order to keep abreast of these developments - we mean functional training approaches which emphasize the link between what trainees learnt with their daily life.

In the other hand we know that the accurate and modern statistical information has become an important part of the basic knowledge and the suitable decisions.

Consequently; all the countries are interested in developing the capacity building with high quality and friendly to provide the necessary data to achieve the suitable and honest indicators to help policy makers to give the right decisions, But to achieve this aim we must interested in education and training because it plays an important role in developing the capacity building and raising the statistical awareness we mean the functional training but the traditional form of education and training (instructor -trainer negative - memorize facts and statistical equations) does not fit in order to create citizen effectively positively contribute at development his society, we do not expect from an individual receive information without any positive role, could be able to express an opinion in any issues that concern the community and solving it, so the researcher finds that functional knowledge is the way to create a statistical person is able to keep pace with developments occurring in the community.

1.1 Definition of Statistics:
There are many definitions of statistics, all covered in the following one of Berenson, et al. (2003);

“The subject of modern statistics encompasses the collection, presentation, and characterization of information to assist in both data analysis and the decision-making process.”
- Statistics can be defined in a more functional as a "science which establishes the concepts we have to live a better life, and develop methods to use information and new discoveries through the methods of data collection and analysis".

This definition is added an excellent addition to making statistics a humanitarian role in life for all members of society.

1.2 Types of statistics (Descriptive and Inferential Statistics)

The study of statistics is usually divided into two categories: descriptive statistics and inferential statistics.

Methods of organising, summarising, and presenting data are called descriptive statistics.

The methods used to determine something about a population on the basis of a sample are called inferential statistics (statistical inference or inductive statistics). Population is the entire set of individuals or objects of interest or the measurements obtained from all individuals or objects of interest. We call a portion, or part, of the population of interest sample.

There are two types of inference approaches in Statistics:
- Classical Inference Approach (Sample information and some non-sample information in form of restrictions).
- Bayesian Inference Approach (Sample information and non-sample information in form of prior distributions).

1.3 Importance of Statistics

In any data-based society, scientists often use large amounts of data to gain a representative sample of the population, and it is not possible to simply look at the numbers and understand what is happening. Statistics allows a trained person to see the significance of data, the relationship between seemingly unrelated phenomena, and predict what may happen in the future or determine what may have happened in the past, and thus, statistics can be seen as the most important tool of the sciences. In this respect, as stated in the NSF 2003 Report, p. 27, “A distinguishing feature of statistics as a discipline is its interaction with the entire spectrum of natural and social sciences and with technology.”

Statistics, however, is not only useful in science. It plays an important role in the planned development of any industry. It can be seen in the business world, in sports, in government, and in any discipline. Decisions regarding social programs, governmental spending, planned industry, and other issues, are made based on statistical analysis. This is because statistics is multidisciplinary in its tools as well as inherently interdisciplinary, the NSF 2003 Report, p. 24 and 60.

2. TRAINING AND STATISTICAL CAPACITY BUILDING

Statistical capacity is a nation’s ability to collect, analyze, and disseminate high-quality data about its population and economy.

WHY is Building Statistical Capacity Important?
Quality statistics are essential for all stages of evidence-base decision-making, including:

- Monitoring social and economic indicators, including progress toward the Millennium Development Goals (MDGs).
- Allocating political representation and government resources
- Guiding private sector investment.
- Informing the international donor community for program design and policy formulation

The statistical training consider as the basic factor in developing the capacity building for the national statistics institute, so we must do our best to use their principle in statistical training to gain trainees statistical awareness and to make statistics meaningful in real life.

2.1 Definition of functional Training

2.1.1 Definition of training

training is the processing which aim to develop the human resources by providing them with knowledge, skillful, increase their capacity building to raise their capabilities and improve their performance and production and achieve the employment goals.

2.1.2 Definition of Functional Training

Functional Training is One of educational approaches that emphasize the link between what trainees learn (statistics, knowledge, facts and information) with their daily life to help them applying what they have studied theoretically on their life, to achieve maximum benefit and to make the material meaningful in real life.

2.2 Importance of the functional training:

- Acquisition trainees statistical information and skills through application of theoretical knowledge.
- Give trainees the nuances of statistical concepts because the trainee in traditional methods may take it without understanding its meaning.
- Developing the capacity for critical thinking and innovation.
- Changing the role of trainer and trainee to more positive interaction in the classroom.
- Help trainees to do his tasks by higher quality.
- Help trainees to achieve himself and increase the confidence to feel that with his distinguish and that he is very important in the institute.
- Functional training is the best way to minimize the wrong ideas and renew and develop the occupation system and achieve the goals of the employment path.

3. THE NATIONAL CENTER FOR STATISTICAL TRAINING (NCST) & RAISING STATISTICAL AWARENESS

According to the CAPMAS Center is established according to the Ministerial Decree no (415) in 1970 organizational structure of the It is regarded as one of the centralized departments related to population statistics and census sector.
3.1 Objectives of the National Center for statistical training:
  • Preparing qualified staff to be able for suitable statistical analysis.
  • Develop the knowledge and the skilful of the staff in the statistical area.
  • Strength the capacity building of the institute and the planning centers in the area of data production and use.
  • Strength the statistical awareness between the public.
  • Support the scientific researches which depend on the country statistics.

3.2 The tasks of the National Center for statistical training
  • Summarize the requirements in the statistical training area.
  • Provide suitable training chances for the data production and users.
  • Training the trainee.
  • Training the university students on the literature research and statistical analysis.
  • Prepare suitable program to increase the statistical awareness for all the population levels.
  • All data of training program should be saved, stored and documented.

3.3 Principles of training
  The training program is implemented in accordance with the following rules:
  1. Theoretical and practical Lectures to explain the foundations of statistical work.
  2. Practical Lectures including the explanation of the scientific basis for the process of collecting data from different sources.
  3. Theoretical and practical lectures for foundations of designing the questionnaire and statistical models, as well as the foundations of revising the audit office and checking the completed forms.
  4. Practical Lectures to explain the concepts, terminologies and instructions of the statistical forms.
  5. Theoretical and practical Lectures for explaining the demographic statistics.

4. TRAINING PROGRAMS AT THE NATIONAL CENTER FOR STATISTICAL TRAINING

4.1 Basic training program:
  • Training program for the university graduates.
  • Developed Training program for university Graduates (follow up).
  • Training program for the intermediate level.
  • Training program for the staff who is working in different statistical places.
  • Summer course for the student who study statistics in the different colleges, with special concentration on the applied statistics.

4.2 Special training program:
  This program only for CAPMAS staff to develop their capacity building, The program including the following subjects:
  • Questionnaire design.
  • Data collection and representation.
Functional training and its impact on the development of statistical awareness

- Statistical measurements.
- Time series.
- Index numbers.
- Samples.
- Modern statistical analysis by using SPSS and Excel.

It is very important to mention that the lecturers are well selected among the distinguish experts with high quality and experiences beside some professors from different universities.

The following table represents the number of the participants from CAPMAS and others between 2000 and 2010 by the type of the program:

### 4.3 The different programs organized by CAPMAS

(The Statistical Training Center)

From 2000 - 2010

<table>
<thead>
<tr>
<th>Type of program</th>
<th>From CAPMAS</th>
<th>Outside CAPMAS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Training program for the university graduates</td>
<td>88</td>
<td>506</td>
</tr>
<tr>
<td>Advanced Training program for university graduate</td>
<td>42</td>
<td>254</td>
</tr>
<tr>
<td>Training program for the intermediate level</td>
<td>90</td>
<td>294</td>
</tr>
<tr>
<td>Training program for the staff who is working in the different statistical places</td>
<td>-</td>
<td>439</td>
</tr>
<tr>
<td>Coverage contracts</td>
<td>91</td>
<td>-</td>
</tr>
<tr>
<td>Ministry of Education</td>
<td>-</td>
<td>22</td>
</tr>
<tr>
<td>University Students</td>
<td>-</td>
<td>104</td>
</tr>
<tr>
<td>Specialized Program</td>
<td>520</td>
<td>-</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>831</strong></td>
<td><strong>1619</strong></td>
</tr>
</tbody>
</table>

### 5. HOW TO DEVELOP STATISTICAL AWARENESS THROUGH FUNCTIONAL TRAINING

There are many methods can used it as a functional tools in teaching statistics such as:

- Interactive Demonstrations
- Presentations
- Practical exercises, individually and in groups
- Discussion
- Use Statistics within a Story
- Playing roles & Brain storm.
- Solving problems
- Cooperative Learning

### 5.1 Do Not Drop Stats Off; Drive Them Home

Many times you will see a speaker go into a list of statistics in order to prove his or her point. Statistics are certainly desirable in speaking, but only if you use them in an effective way. Simply listing statistics is not enough. You must interpret them so your
audience knows exactly what they mean to them. For example, I have often used the following statistic when driving home a point about life-long learning: "Each year you will spend between 500 and 1000 hours in an automobile." This statistic might be interesting, but it certainly is not compelling. That is because it has not been driven home. Here is how you might drive it home:

"Each year you will spend between 500 and 1000 hours in an automobile. If you live to be 75 years old, you will spend approximately 7-10 years of your life in an automobile. Seven years! Here is my question to you. What are you doing with that time? [pause]. Are you simply passing time or using it? Do you know 10 years is enough time for two PhDs? Again, what are you doing with that time? [pause]. I suggest turning your car into a rolling university. You can listen to audio books and quickly become an expert compared to the rest of the country. Oh, but there is one caveat. Please do not use any meditation tapes in the car!"

This is how you can drive a point home by showing your trainees exactly how that statistic affects them and what they can do about it. Do not just drop off the statistic, but drive it home and this core of functional training and we will note that in the following teaching methods.

5.2 Examples for how can be develop the statistical awareness depended on some strategies of teaching

5.2.1 Statistic vs. Story

The problem with simply listing statistics is that they are relatively impersonal and unemotional. Of course we think, "Oh that is terrible" when we hear about millions of people with AIDS or thousands upon thousands of people living in poverty. However, those statistics will not bring us to tears like the story of one person will. A personal story will outdo statistics every time. Please remember that your audience makes decisions based on emotion backed up by logic.

So we can use story to develop statistical awareness about many topics such as Unemployment - Tourism - Immigration – Income and use that at statistical training with a functional way.

5.2.2 Cooperative Learning

An activity involving a small group of learners who work together as a team to solve a problem, complete a task, or accomplish a common goal. (Artzt and Newman 1990).

How to use Cooperative Groups in a Statistical training

It is recommended that the instructor carefully read one of the excellent resources on cooperative or collaborative learning in higher education before incorporating cooperative groups in a statistics class. These resources provide complete information on structuring and monitoring groups, and developing and evaluating group activities. Based on the models of cooperative learning in these references, cooperative group activities for a particular statistics class can be developed. These activities might include:
1. Having groups individually solve a problem and then compare their solutions (e.g., homework problems or problems from the textbook requiring particular skills).

2. Having groups discuss a concept or procedure, or compare different concepts or procedures (e.g., discuss the steps involved in testing an hypothesis, or compare the advantages and disadvantages of using the mean, median, and mode to summarize a data set).

3. Having each group collaborate on a large project involving collecting, analyzing, and interpreting data. Groups may meet in and/or outside of class to work on these projects, and may present the results in a written report and/or an oral in-class presentation.

If Trainees work together but turn in separate reports, these may be rated individually and then a group score based on the average assigned as well. If only a group score is assigned to a group product, students may be asked to volunteer the percentage of their contribution, and that may be used to determine their share of the group points. Or, a group score may be assigned and everyone receives that score. Methods will vary based on the types of projects and students.

**Example: Measures of Center**

1. Each group of Trainees is given some different data sets (e.g., prices of running shoes, fat content of fast foods, Olympic medals, temperatures for a month).

2. In your group, discuss each of the three measures of center and make sure that everyone understands what each measure is and how it is calculated.

3. Discuss the advantages and disadvantages of using each of the three measures to summarize a data set.

4. For each of the distributed data sets, determine which measure of center would be most appropriate as a single number summary and why.

5. Turn in one written summary of your discussion. Be sure to include a description of each measure and how it is calculated, advantages and disadvantages of each measure, and a discussion of which measure of center is most appropriate to use in representing each data set and why.

**5.2.3 Using Computer in statistical training**

Statistical awareness can be developed through some software programs such as the toolkit, a program for the conservation and documentation of data surveys statistical and thus can develop awareness of the trainees some statistical concepts such as sampling procedure - the response rate -weighting - research community and also giving them awareness of the means of data collection and the stages of statistical research by assigning them one documented statistical surveys using the toolkit.
Note: researcher do a guide to how to use the Toolkit has not been published yet.

5.2.4 Use statistical programs (SPSS program) as an educational tool

Example (1) (correlation) datasets of this example we can find it in samples of SPSS program (car_sales)

Director of the company wants to evaluate the performance of sales in his company and is there a relationship between the improvements made to the fuel efficient cars, sales of the company.

Table (1) Relation between fuel efficient cars, sales of the company

<table>
<thead>
<tr>
<th>Correlations</th>
<th>Sales in thousands</th>
<th>Fuel efficiency</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sales in thousands</td>
<td>Pearson Correlation</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Sig. (2-tailed)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>N</td>
<td>157</td>
</tr>
<tr>
<td>Fuel efficiency</td>
<td>Pearson Correlation</td>
<td>-.017-</td>
</tr>
<tr>
<td></td>
<td>Sig. (2-tailed)</td>
<td>.837</td>
</tr>
<tr>
<td></td>
<td>N</td>
<td>154</td>
</tr>
</tbody>
</table>

Table above shows that the relationship is Negative Correlation and not significant, that mean fuel efficiency has no impact in sales, but it must be emphasized to the trainees should see the previous result in-depth consideration to the data that we are to reach the result indicates the awareness of statistical.

For example, if we know that there are two types of cars (Automobile-truck), even if we did we split file on the basis of vehicle type and repeat the correlation we will show the following result.
Table (2): Correlation between fuel efficient cars, sales of the company after split file on the basis of vehicle type

<table>
<thead>
<tr>
<th>Vehicle type</th>
<th>Sales in thousands</th>
<th>Fuel efficiency</th>
<th>Pearson Correlation</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>Automobile</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sales in thousands</td>
<td>Pearson Correlation</td>
<td>1</td>
<td>.262**</td>
<td>116</td>
</tr>
<tr>
<td>Fuel efficiency</td>
<td>Pearson Correlation</td>
<td>.262**</td>
<td>1</td>
<td>114</td>
</tr>
<tr>
<td>Truck</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sales in thousands</td>
<td>Pearson Correlation</td>
<td>1</td>
<td>.011</td>
<td>41</td>
</tr>
<tr>
<td>Fuel efficiency</td>
<td>Pearson Correlation</td>
<td>.011</td>
<td>1</td>
<td>40</td>
</tr>
</tbody>
</table>

**Correlation is significant at the 0.01 level (2-tailed).

The previous table indicates that spending on improving fuel efficiency for Automobile sales affects This is the statistical awareness that we want to lend it to the trainees.

6. FUNCTIONAL TRAINING FOR FIELD RESEARCHERS

A field researcher Must know that there are differences and diversity of personalities and behaviors of respondents in a way to answer questions posed in the search form, and therefore must deal with all this variation to reduce the errors of response and that arise from giving the respondent data is inaccurate or did not express the data to be obtained. Here comes the role of functional training in order to give researchers this awareness to get the statistical data of high quality.

Table(3): Types of respondents and how to deal with them

<table>
<thead>
<tr>
<th>Ser</th>
<th>Type of responder</th>
<th>How to deal</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Respondent Silent: Is a person who refuses to speak or respond to the researcher, not showing any impressions of either acceptance or rejection, and thus difficult to know what is going on in his mind.</td>
<td>need this kind to attempt to break the silence has questions include yes or no at the beginning to talk to him, so the researcher can gain confidence and encourage them to break the silence and cooperation. - a field researcher must have patience and to gain the confidence of the skin under examination, and asked his opinion frankly as asking him questions</td>
</tr>
<tr>
<td>2</td>
<td>Respondent Jabber: Usually wish this Category talk all the time without the sense or meaning and nature of debate and discussion and enthusiasm and is not logical in how to respond to the question</td>
<td>need this Category to the special package so that the researcher can ask his wish to speak to the benefit of research, conducted by. - the investigator must give him the attention and tries to attract the main theme in a friendly way. *- Warning Do not try to silence effect or Jabber force.</td>
</tr>
<tr>
<td>Ser</td>
<td>Type of responder</td>
<td>How to deal</td>
</tr>
<tr>
<td>-----</td>
<td>----------------------------------------------------------------------------------</td>
<td>-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
</tbody>
</table>
| 3   | Respondent Sceptic: Is a person who does not easily believe what he says always have researcher and critic may be tactless | - This Category always needs to provide evidence on the proposal of the researcher.  
- You must show him and prove his friendship researcher safety purpose.  
- is necessary to emphasize that the data has to be delivered will be kept confidential and will not be used for purposes other than research |
| 4   | Respondent nervous: Usually a Category has a nervous reaction, a fast of the most difficult types of respondents to the inability of the researcher pursuing his ideas and rapid transition from topic to another. | - must be removed from the details as much as possible  
- you should try to return to his cool through simply talking to him and help him arrange his ideas to serve the objectives of the research.  
- Trying to answer his questions honestly and deal with objections on the basis of logical |
| 5   | Respondent with special needs: May be dealing with Quested has a disability be certain if the loss of one of their senses, or has difficulty speaking, this is a Category difficulty in dealing with the researcher. | - must not be exaggerated to express the idea while trying to transfer him to not try to quit and left undisturbed for an interview and not the multiplication of detailed questions that may be embarrassed, should also be wary of dealing with him on the basis of compassion, because that will make him feel weakness. |

7. MEASURE THE SATISFACTION OF TRAINEES

**Title of study**

Measure the degree of satisfaction of the trainee for the training program (SPSS).

**7.1 Methodology of the study:**

Researcher used the descriptive analytical method have been found on some of the books - Literature Internet-related subject of the study (Likert scale and trends).

**7.2 Community study:**

A sample consists of (40) trainee from trainees enrolled in the training program (SPSS) at (Statistical Training Center), the instrument of the study were distributed to them (survey form) in order to identify the degree of satisfaction about the training program.

**7.3 Tool of the study:**

A questionnaire has been prepared by the researcher(*)

The questionnaire has been divided into three Sections as follows:

- First Sections (the objectives of the training program):
- Discusses the extent to which the training program with work and raise the capabilities of the trainees on the statistical analysis and dealing with the program SPSS

(*)questionnaire about the degree of satisfaction of the trainee.
• Second Section: (Training on the program SPSS)
  • Discusses the training program of the SPSS (the timetable - the approach adopted in training - the view from the instructor.
  • Section third: (advantage of the program of the SPSS)

Discusses the extent to which trainees apply what they have learned in their work and how they gain skills of statistical analysis on a program of the SPSS.

7.4 Procedures

- Do some statistical treatments using the SPSS program with a view to the results of the survey has been the statistical treatments are as follows:
  • frequency tables and graphs of demographic variables (gender - age - educational status - employment status - Place of Work).
  • Calculate validity and reliability of the survey questions using a factor (Cronbach alpha Cronbach's Alpha).
  • Calculate trend of the trainees in accordance with the methodology for the Likert and using the SPSS
- Calculate correlation between the three sections of the questionnaire (the objective of the training - implementation of the training - advantage of the training program).

7.5 Results

1. sample, consists of 40 people (20 male / 20 female) were distributed according to the state functional is (50% overall contract - 22.5% held temporary - 27.5% specific) and for the sectors is 10% the presidency of CAPMAS - 17.5 % sector of economic statistics and tactical - 17.5% population statistics and censuses - 12.5% Sector Regional Branches - 25% of the information technology sector - 17.5% from outside the system
2. Calculated direction of the trainees and the extent of satisfaction with the training program, using Likert scale has been reached following results:
  • the direction of the trainees about the training program in terms of the goal of it was positive, although it was noted that there are neutral to the goal of "upgrading the skills of statistical analysis of the trainees, using SPSS program and it is due from a working group to that the skills of statistical analysis is not understandable for workers in other sectors that is not use statistics "that for the first section (the objective of training) the following figure showed this results.
Fig. (2): Trend of the trainees about (objective of training)

- Results showed that trend of trainees is positive about this section II with the exception of a degree neutral with respect to the timetable, although it was sufficient or not, because from the perspective of researcher that the trainees when applied to them the questionnaire was not training program is over, which made neutral respond the following figure showed this results.

Fig. (3): Trend of the trainees about ((training program))

- The trainees showed a positive trend is also about the third section. But noted that there is neutrality on the part of some trainees to express their opinion on the point "I can apply what you were trained in the business which I do" because from the viewpoint of the researcher that there are some trainees scope of their work is far from the statistics. The following figure showed this results.
There is a correlation is weak and not significant between (objective of the training and implementation of training), where the value of the correlation coefficient = .099, non-significant (Sig. (2-tailed) = .544) indicating from the point of the Researcher to the lack of clarity of purpose of the training for trainees. The following table showed this result.

**Table (4): correlation between the total of the three section (the objective of the training / training program / benefit)**

<table>
<thead>
<tr>
<th></th>
<th>Total Average answers</th>
<th>Mean Total answers</th>
<th>Mean Total answers</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>The first section</td>
<td>The second section</td>
<td>The third section</td>
</tr>
<tr>
<td></td>
<td>(The objective of</td>
<td>(training program)</td>
<td>(to take advantage</td>
</tr>
<tr>
<td></td>
<td>training)</td>
<td></td>
<td>of the training</td>
</tr>
<tr>
<td>Total Average</td>
<td>Pearson Correlation</td>
<td></td>
<td></td>
</tr>
<tr>
<td>answers</td>
<td>1</td>
<td>.099</td>
<td>.607**</td>
</tr>
<tr>
<td>The first section</td>
<td>Sig. (2-tailed)</td>
<td>.544</td>
<td>.000</td>
</tr>
<tr>
<td>(The objective of</td>
<td>N</td>
<td></td>
<td></td>
</tr>
<tr>
<td>training)</td>
<td>40</td>
<td>40</td>
<td>40</td>
</tr>
<tr>
<td>Mean Total answers</td>
<td>Pearson Correlation</td>
<td>.099</td>
<td>1</td>
</tr>
<tr>
<td>The second section</td>
<td>Sig. (2-tailed)</td>
<td>.544</td>
<td>.658</td>
</tr>
<tr>
<td>(training program)</td>
<td>N</td>
<td>40</td>
<td>40</td>
</tr>
<tr>
<td>Mean Total answers</td>
<td>Pearson Correlation</td>
<td>.607**</td>
<td>.072</td>
</tr>
<tr>
<td>The third section</td>
<td>Sig. (2-tailed)</td>
<td>.000</td>
<td>.658</td>
</tr>
<tr>
<td>(to take advantage</td>
<td>N</td>
<td>40</td>
<td>40</td>
</tr>
<tr>
<td>of the training</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>program)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
8. RECOMMENDATION

- Establishment of lectures and seminars statistical regularly and continuously for spread of statistical awareness among citizens and understand the concept of statistics.
- Training the staff on the latest technologies and software and functional statistical methods.
- Try to reduce the time period for the issuance of bulletins and in accordance with the timetable.
- Strengthen the position of the National Statistical Office on the Internet and updated on an ongoing basis
- Receiving student delegations from universities and private schools, government and inform them of the activities of the National Statistical Office in the field of statistical data and dissemination, and provide them with some of the releases.
- Giving lectures in schools and institutions aimed at introducing the National Statistical Office and contribute to the further statistical awareness.
- Distribution of a special form designed to solicit the views of users of data on the degree of satisfaction with the device and press the accuracy and quality of data as part of the process of feedback on the work of the device.
- Investment and various media audio-visual and print investment represent for the dissemination of statistical awareness through seminars statistical live "forums, advertising," and not only the periods of the implementation of the census year with the necessity of activating the Media Research statistics also to identify the respondents view each search before starting implementation.
- A guide to regular publications and statistical surveys, which the National Statistical Office in each year so as to be distributed in public places in order to obtain all the studies and research with clarifying the importance of statistical work and its aim is.
- Use the style of advertising on the statistical means of transport of CAPMAS and the motto was, for example, "right statistics equals a better life."
- Teaching subject under the name "statistical awareness" to include some research in various fields, which show the importance of statistics in life
- Teaching how to use a methodology for the Likert has been prepared scientific material necessary to do so.
- Emphasize the objectives of the training courses at the advertised
- Using the functional methods in teaching statistics such as current events - the story - collaborative learning - brainstorming ... etc. is very important because it gives meaning for using statistics to face the problems of society.
- Interest in measuring the satisfaction of the trainee after the completion of statistical training programs.

CONCLUSION

Based on the above, no one can deny that Statistics plays an important role in our modern life, where became familiar to us, and represents an important aspect of the information which we read every day, such as tables of points achieved by the football clubs which we read it in newspapers, magazines, and estimates of the forecast weather and stock indexes and the government's achievements in the field of housing and reconstruction and changes in the exchange rates and commodity prices.
Therefore we must development of statistical awareness through functional training and his ways such as cooperative learning, the results of statistical surveys, use computer to made statistics a functional role in our life and in solving our problems in a scientific way.

REFERENCES

A PROPOSED APPROACH ON MULTIVARIATE CLUSTERING ANALYSIS USING SIMULATED ANNEALING – A CASE STUDY ON LOCATION PROBLEM

Mahdi Bashiri¹ and Mohammad Khanloo¹
¹Department of Industrial Engineering, Shahed University, Tehran, Iran
Email: bashiri.m@gmail.com, khanloo.mohammad@gmail.com

ABSTRACT

Cluster analysis is a useful technique in multivariate statistical analysis. Hierarchical cluster analysis and K-means have been used for data analysis in previous studies. However, the K-means algorithm can be improved by using some meta-heuristics algorithms. In this study we propose simulated annealing based algorithm for K-means in the clustering analysis which we call it as SA-K-means. To show the effectiveness of the proposed approach, some numerical examples of location problems have been tested by the proposed approach. The results show that the performance criteria of the proposed approach including sum of squares between the elements within each cluster (SSw), sum of squares between clusters (SSb) and also the lambda value (\( \lambda = SS_b / SS_w \)) are satisfactory.

KEYWORDS

Multivariate Clustering Analysis, K-means, Simulated Annealing, Sum of Squares, Location Problem.

1. INTRODUCTION

With the growing nature of data in the daily business environment, the analysis and implementation of data seems to be a basic task of management. Data mining is a useful and efficient process of analyzing such data and clustering is a popular data analysis and data mining technique. The term clustering is used for different data type especially when the subject of study is about speech or image data. For speech and image data the term “vector quantization” is used instead. Although they seem to be as separate problems, both the content and solution styles are the same. Clustering procedures partition a set of objects into clusters such that objects in the same cluster are more similar to each other than objects in different clusters according to some predefined criteria. The problem can be generalized as follows: given a set \( n \) data points in \( d \) dimensional space, we must determine how to assign a set \( C \) of \( c \) points, called centers, in \( n \) so as to optimize based on some criterion. In most cases, it is natural to assume that \( n \) is much greater than \( c \) and \( d \) is relatively small. This formulation is an example of unsupervised learning. The system will create groupings based only the criterion and the information contained in the \( n \) data point. K-means is a popular algorithm that was first presented over three decades ago. The mentioned criterion minimizes the total mean-squared distance from each point in \( n \) to that point’s closest center in \( c \). Data clustering, which is an NP-complete problem [1] of finding groups in heterogeneous data by minimizing some measure of dissimilarity, is
one of the fundamental tools in data mining, machine learning and pattern classification solutions. It is the process of grouping similar data items together based on some measure of similarity (or dissimilarity) between items.

Clustering is a component of exploratory data analysis and is useful for generating hypotheses about data.

The standard clustering process consists of the following steps:

1. Data preparation and attribute selection,
2. Similarity measure selection,
3. Algorithm and parameter selection,
4. Cluster analysis and
5. Validation.

In this paper, the dependence of the K-means algorithm is very strong initial response and this makes algorithm in most cases fall into the trap of local optimization. Also, it has been proven that the metaheuristic algorithms are very good performance and high speed to reach response in location problems. Therefore, we have used SA algorithm in this paper. Several numerical examples we have solved using the proposed algorithm and we compared the answers obtained from the proposed and K-means algorithms, as noted in the previous section, the lambda value is validation criteria, therefore the definition of lambda, the lambda value is much higher clustering is done carefully. As numerical examples in the Table 2 to see all of the lambda value that obtained from proposed algorithm is larger than them that obtained from K-means. So, the proposed algorithm has very good performance in the location allocation problem.

In the Table 1 we review some papers that they used clustering methods for location-allocation problems. But they didn’t use metaheuristic algorithm for this problem.

<table>
<thead>
<tr>
<th>Authors</th>
<th>Problem titles</th>
<th>Clustering method (Technique for solving)</th>
<th>Metaheuristic algorithms</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hsieh and Tien [2]</td>
<td>Location-Allocation</td>
<td>Self organizing maps</td>
<td>-</td>
</tr>
<tr>
<td>Bischoff and Dachert [3]</td>
<td>Location -Allocation</td>
<td>K-means</td>
<td>-</td>
</tr>
<tr>
<td>Murat et al. [4]</td>
<td>location–allocation problem with dense demand</td>
<td>K-means with heuristic algorithm</td>
<td>-</td>
</tr>
<tr>
<td>Correa et al. [5]</td>
<td>maximal covering location-allocation problem</td>
<td>K-means</td>
<td>-</td>
</tr>
</tbody>
</table>
As noted above, they have not used metaheuristic algorithms for location allocation problem, given the good performance of metaheuristic algorithm; we have used the SA algorithm in this paper.

The remainder of this paper has been organized as following; in the next section the proposed algorithm based on the simulated annealing for multivariate clustering analysis (SA-K-means) will be discussed. The proposed algorithm has been tested with several hypothesis numerical examples which have been reported in section 3. Finally the conclusion has been mentioned in the last section.

2. PROPOSED ALGORITHM (SA-K-means)

Location-allocation (LA) problem is to locate a set of new facilities such that the transportation cost from facilities to customers is minimized and an optimal number of facilities have to be placed in an area of interest in order to satisfy the customer demand. This problem occurs in many practical settings where facilities provide homogeneous services such as the determination and location of warehouses, distribution centers, communication centers and production facilities. The basic components of location allocation problems can be thought to consist of facilities, locations, and customers. In the following fig, an example for location allocation problem with c=3 and some of customers that served with facilities. In this fig symbol “▲” showed facility location and “◇” showed customers.

As seen in previous sections one of the most important problems in center-based clustering is convergence to the local optimum. For combinatorial optimization problems there are some heuristic methods to find the global optimum point instead of local one. Tabu search [6], simulated annealing [7], evolutionary methods, swarm optimization methods are examples of these heuristic methods. In order to overcome the local optima problem lots of studies have been done in clustering and vector quantization literature.
A proposed approach on multivariate clustering analysis…

Hansen and Mladenovic [8] proposed a method, J-means, based on local search heuristic. Zulal Gungor and Alper Unler [9] proposed K-harmonic means data clustering with simulated annealing heuristic. Frnti et al. [10,11] proposed randomized local search, Cano et al. [12] proposed greedy randomized adaptive search procedure (GRASP) and Kanade and Hall [13] proposed an ant colony optimization method to solve local optima problem in clustering. Some authors [14,15,16,10,17] used Tabu search method, while some others [18,19] using simulated annealing, and some other [20,21] using hybrids like genetic simulated annealing or tabu search with simulated annealing, to overcome this problem. The main idea of their algorithms is to use some heuristics (Tabu search, simulated annealing, random search etc.) to generate non-local moves for the cluster centers and to select the suitable best solution. All of the algorithms mentioned above use either K-means for the clustering problems.

In the clustering methods, the validation criterion is the ratio of between clusters sum of square and each cluster sum of square. Mentioned criteria can be computed according to the following equation:

\[ \lambda = \frac{SS_b}{SS_w} \]  

(1)

Simulated annealing (SA) algorithm needs an initial solution to start. In this study, we can give initial solution in two forms to the SA algorithm. The first one is the final K-means algorithm solution and the other is the random generated solution. If we take the initial solution from K-means algorithm, final computed \( \lambda \) value and cluster that get from K-means algorithm as initial solution for SA, but, if random initial solution for clustering in SA is used, initial \( \lambda \) value is zero. SA operation in two form is same, therefore, algorithm begin to change items into the different clusters. If new \( \lambda \) value was major than before \( \lambda \) values, this value is new best \( \lambda \) value, but if new \( \lambda \) value smaller than before \( \lambda \) value, so, we accept it by probability of description in SA, therefore, SA-K-means is not convergence to local optima. Finally, SA-K-means display maximum \( \lambda \) value and optima cluster that relation to maximum \( \lambda \) value. Since, in this study, we descript \( \lambda \) value for clustering validation. Therefore, if \( \lambda \) value increase, we can take a result that the clustering has high accuracy. After compare results of SA-K-means and K-means algorithm on several case studies in location problem (Location Allocation problem), we take a result that SA-K-means make higher \( \lambda \) and accuracy. In Fig. 2 we presented pseudo code for proposed algorithm. Table of \( \lambda \) values is in numerical analysis section.
Begin
- Input $c$ (number of cluster) and matrix $x$ (customers location)
- Choose $c$ as initial centers
  \[
  \text{final solution of K – means} \\
  \text{Randomly}
  \]
- Set $T=T_{\text{max}}$
- Set $T_0$
- Set cooling schema ($r$)
\[
\begin{align*}
\text{Repeat:} & \\
\text{Repeat:} & \\
& \text{Change the allocation of observations in to the clusters} \\
& \text{(Neighborhood solution)} \\
& \text{Calculate the clustering index ($\lambda$-value)} \\
& \text{If } \Delta\lambda > 0 \text{ then swap old and new solution} \\
& \text{Else} \\
& \text{Generate a random number ($s$)} \\
& \text{If } e^{-\Delta\lambda/T} > s \text{ then swap old and new solution} \\
& \text{End if} \\
& \text{End if} \\
& \text{Until maximum iteration} \\
& \text{Until } (T=T_0) \text{ stopping condition}
\end{align*}
\]
End

Fig. 2: Pseudo code for proposed algorithm

3. NUMERICAL ANALYSIS

Since in this study, we used $\lambda$ value for clustering validation, by refer to Table 2, you can see that we make five clustering operation for all of them. Numerical examples related to the location allocation problem, as in them, $n$ is the number of customers that to be serve by facilities and $c$ is the number of facilities that must be locate. Numbers into the Table 2 is $\lambda$ value for every numerical example. The numbers of every numerical examples and all of the numbers base on $n/c$ in two groups, that if $n/c$ value is major then $\lambda$ value is increase too, so, $\lambda$ and $n/c$ values is direct relation. SA-K-means by two way for initial solution (K-means and random) is same together and both of them has high accuracy. It should be noted, column corresponds to the K-means, $\lambda$ values that calculated from the K-means algorithm and column correspond to the SA-K-means, which is includes two sections of this initial solution for SA-K-means, K-means (initial solution) column, column in which the initial solution for SA-K-means is the final solution that obtained from K-means and random (initial solution) column, column in which the initial solution for SA-K-means is randomly. As mentioned, whatever lambda value is more, the clustering has higher accuracy, so, location and allocation well done.
Consider a problem which has $n=33$ (number of customers) and $c=4$ (number of facility). Fig. 1 show that the clustering of this problem by proposed algorithm, in this fig symbol “$\triangle$” show that facility location and other symbol show every cluster in this problem. For this problem the initial solution for SA-K-means algorithm is final solution that obtained from K-means algorithm.
4. COMMENTS AND CONCLUSION

Clustering is an efficient way of reaching information from raw data and K-means is a basic method for it. Although it is easy to implement and understand, K-means has serious drawbacks. Over the past 30 years, many papers have presented ways in which to improve K-means. Some focus on creating good initialization methods, while other look at finding optimal values for K, and some others for finding globally optimal solutions. In this study we have proposed a K-means data clustering algorithm with simulated annealing heuristic which we call SA-K-means. The algorithm has been implemented and tested on several well-known numerical examples in location allocation problem. We have presented with the experiments that our algorithm (SA-K-means) outperforms to K-means, and SA-K-means from the point of performance value in most cases. On the other hand, SA-K-means has high accuracy. In the future we are planning to study K-means data clustering with other heuristic based search methods like swarm optimization, genetic algorithms or some others. And also we are planning to study with real life domains and to show our algorithm will work well with those domains.

REFERENCES

STATISTICAL ANALYSIS OF THE NUMBER OF THE GALAXIES IN CUBIC CELLS IN THE UNIVERSE

Fahimeh Mostajeran
Department of statistics, Shahid Behesti, Tehran, Iran
Email: fahimeh_mostajeran@yahoo.com

ABSTRACT

Today, modern astronomers extensively study about the distribution of the galaxies in the universe. Part of these research includes statistical analysis of the number of the galaxies in cubic cells (cell counts), since working with such a data is easier than raw spatial data with their complex spatial patterns. There exist various models for cell count distribution (Martínez and Saar, 2002). Classical point process statistics would probably suggest negative binomial distribution because of the high degree of clustering of galaxies. This model has also been used by cosmologist (Elizalde and Gaztañaga, 1992 and Betancort-Rijo, 2000). Recently, discrete Weibull distribution was fitted to data (Ghorbani et al., 2006). This approach seems to be appropriate as there are some models for the galaxy clustering process that are based on the fact that the universe is seen by cosmologists as a result of a growth process (Bernard et al., 2004). Furthermore, since the number of empty cells is significantly greater than others, Zero-Inflated distributions is also suggested as a new alternative. Also, Zero-Inflated weibull and Zero-Inflated Negative Binomial are introduced to fit to data. In this paper, after studying the dispersion of the galaxies in our real studied space, we fit all proposed distributions to data by numerical methods and then the goodness-of-fit of these models have been compared by the Akaike Information Criterion (AIC).

KEY WORDS

Spatial statistics, cell count distribution, zero-inflated distribution, discrete Weibull, growth models, point process

1. INTRODUCTION

The last half of the twentieth century saw cosmology develop into a very active and diverse field of science. This was largely due to the development of observational techniques that allowed astronomers to observe extremely distant regions of space. This motivated a flow of new theories about the evolution of our universe and the formation of the large-scale structure we found in it. The main tools to compare theoretical results with observations in astronomy are statistical, so the new theories and observations also initiated an active use of spatial statistics in cosmology.

Many of the statistical methods used in the analysis of the large-scale distribution of matter in the universe have been developed by cosmologists and are not too rigorous. In many cases, similar methods, sometimes under different names, had been used for years in mainstream spatial statistics. In the late 1950s, when the Berkeley statisticians J.
Neyman and E. Scott carried out an intensive program for the analysis of galaxy catalogs, the connection between spatial statisticians and cosmologists was a fruitful one. However, in the following 30 years cosmologists were not, in general, aware of developments in statistics, and vice versa. Fortunately, recent years have brought the resumption of a dialog between astronomers and mathematicians. We hope that this dialog will continue and will be useful. Cosmology is a good field for applications of spatial statistics. Its well-defined and growing data sets represent an important challenge for the statistical analysis, and therefore for the mathematical community (Martinez, V.J. and Saar, E., 2002).

The present paper describes some spatial statistics methods to find the distribution of the number of galaxies in cubic cell as a tool to study the complex point pattern of galaxies in the universe.

2. GALAXIES AND GALAXY CLUSTERS

Galaxies are a set of millions of stars and star clusters along with gas, dust and a lot of invisible substances. Almost all of us have observed a foggy white band in the sky at a dark night. This white band that is known as Milky Way is indeed consists of millions of stars which seem dim due to far distant. Our sun family is settled at one corner of this great set and is just a dust against its greatness. Diameter of the galaxy is about one half of the distance of the earth to the moon in a scale that we assume the sun equal to a pin.

Other galaxies have been discovered after the Milky Way galaxy and also it has been specified that objects like small and big Magellanic Clouds and Andromeda galaxy are full of stars which are assumed to be nebular. Now we know that the world is composed of millions of galaxies. These galaxies are divided into three groups in terms of shape: elliptical, spiral and irregular. Emergence of the galaxies and reason of their different shapes are a hot topic among astronomers and no convincing idea has still been represented in this regard.

Galaxies are often in group form and the number of single galaxies is few, as number of stars that do not belong to a galaxy is very insignificant. Clusters may include two to thousands of galaxies. Our galaxy is inside of one of these clusters that are known as the local group. Astrological observations from depth of the space show that galactic clusters are inside of larger clusters which are called super cluster or cluster of clusters (Mosadegh, 2003).

Position of galaxies in the space is a random variable. Finding a random complicated pattern which models their place (spatial point model) and responding to the question that whether the settlement position is in cluster, regular or random form (that is realization of a stationary Poisson point process) or not are very important.

Therefore, one of the important applications of point processes is in issues related to cosmology and finding distribution of the number of galaxies in cubic cells. In such issues a region of the space is considered with very large and distinct sizes and is divided into regions with smaller sizes (cells); then the number of galaxies in these regions is counted. Obviously this number is a random variable and the purpose is to find an
appropriate distribution for this variable to be able to model dispersion manner of galaxies in the primary region by help of it.

3. TESTING THE HYPOTHESIS THAT A POINT PROCESS IS POISSON

In order to respond to some issues regarding the spatial pattern of observations including whether places are clustered spatially or are inclined to regular distribution (Poisson point process) some methods are used that their purpose is to convert and summarize spatial data into helpful descriptive statistics. These could be used to explain models that may be fitted to a complicated point pattern.

One of these methods that has been used in this paper is quadrate count method in which the intended region in \( \mathbb{R}^2 \) or \( \mathbb{R}^3 \) is divided respectively into sub regions with equal area or volume and the number of events inside each sub region is counted (see Cressie, 1993).

3.1 Index of dispersion test

Index of dispersion test is designed based on quadrate count method. First the intended region \( W \) is divided into \( k \) sub regions with equal areas for example small squares. Then, the number of points in each sub region i.e. \( v_i \) are recorded. Now if we have Poisson field, \( v_i \) s are independent and co-distribution and average number of points in each sub region is equal to \( \frac{\lambda A(W)}{k} > 1 \). Difference of \( v_i \) s with this amount or a behavior that is very monotonous shows deviation from Poisson assumption. Accordingly, statistic of index of dispersion test is:

\[
I = \frac{(k-1)s^2_v}{\bar{v}}
\]

where \( k \) is the number of sub regions, \( s^2_v \) is variance of the number of points in the sub regions and \( \bar{v} \) is average number of points in each sub region that is equal to \( \frac{N(W)}{k} = \frac{n}{k} \).

If \( I > \chi^2_{(k-1,1-\alpha)} \) or \( I < \chi^2_{(k-1,\alpha)} \), Poisson assumption is rejected ( \( \alpha \) is type I error).

Also we must have \( k>6 \) and \( \frac{\lambda A(W)}{k} > 1 \). If Poisson assumption is rejected in point model, we can assume in the first position that points have cluster state and in the second position it means that points are distributed more regularly than the usual stationary Poisson point field (see Cressie, 1993).

3.2 Dispersion Indices

In a spatial analysis when completely randomness hypothesis is rejected, the next step is measuring departure value of completely randomness. Some of Indices are
Statistical analysis of the number of the galaxies in cubic cells in the universe

<table>
<thead>
<tr>
<th>Reference</th>
<th>Estimator</th>
<th>Index</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fisher (1922)</td>
<td>$\frac{S^2}{\bar{X}}$</td>
<td>I</td>
</tr>
<tr>
<td>David and Moore(1954)</td>
<td>$\frac{S^2}{\bar{X}} - 1$</td>
<td>ICS</td>
</tr>
<tr>
<td>Duglas (1975)</td>
<td>$\frac{\bar{X}^2}{S^2 - \bar{X}}$</td>
<td>ICF</td>
</tr>
</tbody>
</table>

Indices I and ICS are based on the equality of mean and variance of Poisson distribution. Barteko(1968) showed that while the number of points in sub region has Poisson distribution, the expected value for ICS is zero. The values greater than 1 for I and greater than zero for ICS show clustering of points (events in the space). If the value of ICS is less than zero, it shows that the intended space is regular.

$$ICF = \frac{\bar{X}^2}{ICS}$$ is the average number of clusters in each sub region(Cressie, 1993).

4. VOID PROBABILITY FUNCTION

Recent results concerning the galaxy distribution at scales<100h-1 Mpc show a number of characteristics which cannot be described by conventional statistical models. There is clearly a need for new kind of statistical models and statistical indicators, among those I wish to emphasize the advantages of the void probability function (VPF), and its particular convenience for studying the galaxy distribution.

For a given sample of galaxies, the VPF is defined as the probability that a randomly selected volume, of specified size V and shape contains no galaxy. This quantity has many advantages in cosmological research. It can calculate from dynamical or statistical models. It can be predicted from the set of all correlation functions or from the set of count probabilities, for which it acts as a generators, see Lachièze-Rey (1987).

The VPF is basically a function of two variables the density of sample and the size V of the volume. Here we try to find and suggest distributions which estimate void probability well. However, the VPF dependence on sample size and volume size needs more researches which are not in the scope of this paper.

5. CELL COUNT DISTRIBUTIONS

As mentioned above, the number of events in each sub region is a random variable. The aim is finding an appropriate distribution to fit it. For example, under a completely spatial random model, the number of points in each sub region A1 with area (volume) $|A1|$ has Poisson distribution with parameter $\lambda|A1|$. In present paper, after checking the hypothesis that a point process is Poisson, the departure measurement of completely randomness is calculated. In the following, I try to find proper model for the number of points in each sub region. All applied distributions are introduced in this section.
5.1 Negative Binomial Distribution

The Negative Binomial is one of the most frequently used discrete distributions in statistical literature especially in modeling count data with over-dispersion (variance is greater than mean). In classical point process statistics, it is proposed to fit the number of events in disjoint sub regions. As a suggested model, this distribution is applied to fit the cell count distribution of galaxies.

If random variable $X$ has Poisson distribution with parameter $\lambda$ and $\lambda$, the mean of distribution, has Gamma distribution with parameter $a$ and $k$, then the probability mass function of $X$ is:

$$P(X = x) = \frac{\Gamma(k + x)\lambda^k}{x!\Gamma(k)(1 + \lambda)^{k+x}}, k = 0, 1, 2, \ldots, a, k > 0$$ (1)

The mean and the variance of Negative Binomial distribution are $E(X) = \frac{k}{a}$ and $E(X) = \frac{k}{a}$, respectively.

5.2 Weibull Distribution

As an alternative, the Weibull distribution is suggested to fit the cell count distribution of galaxies. Nakagawa and Osaki (1975) obtained Discrete Weibull distribution from its continuous version. A random variable $X$ with the distribution function

$$F(x) = 1 - e^{-\lambda x^\beta}, x > 0$$

is said to have a Weibull distribution with parameters $\lambda$ and $\beta$. The probability mass function and the first and second moments of discrete version of Weibull distribution are

$$P(X = x) = e^{-\lambda(x-1)^\beta} - e^{-\lambda x^\beta}, x = 1, 2, \ldots, \lambda, \beta > 0$$

$$E(X) = \sum_{x=0}^{\infty} e^{-\lambda x^\beta}$$

$$E(X^2) = 2 \sum_{x=0}^{\infty} xe^{-\lambda x^\beta} + E(X)$$

In this paper, the parameter $\beta$ is replaced by the real-valued $\beta = 1/m$ and $(x-1)$ is replaced by $(x-1) = k$. The Discrete Weibull distribution is written as

$$P(X = x) = e^{-\lambda(x-1)^\beta} - e^{-\lambda x^\beta}, x = 1, 2, \ldots, \lambda, \beta > 0$$ (2)

The approach of using Weibull distribution to fit the cell count distribution of galaxies is based on stochastic growth models. Consider a random variable $X(T)$ which is constructed as follows. There is a deterministic law of growth $X(t)$ according to which some size or number variable increases in time $t$. This growth process is stopped at a random time $T$ so that the growth process ends with the value $X(T)$. Thus $X(T)$ is simply
a function of the random variable T. The value of this approach of modeling lies in the possibility of physical interpretation of the parameters of the various distributions.

This approach seems to be appropriate as there are some models for the galaxy clustering process that are based on the fact that the universe is seen by cosmologists as a result of a growth process, see Bernard et al. (2004).

If $T \sim \text{Exp}(\lambda)$ then the random variable

$$X = T^m$$

has a Weibull distribution with parameters $\lambda$ and $1/m$. Note that this is not the so-called three-parameter Weibull distribution. However, as for many of our cells the number of galaxies is small, we do not work with the continuous Weibull distribution but instead by its discrete version with PMF (2), see Ghorbani et al. (2006).

### 5.3 Zero-Inflated Distributions

Recently, zero-modified discrete distributions called Zero-Inflated parameter (IGPSD) are used to model counts that encounter disproportionally large frequencies of zeros. Let $X$ be an arbitrary nonnegative integer-valued random variable such that

$$P(X=j)=p_j, \quad j=0, 1, \ldots, \quad \sum_{j=0}^{\infty} p_j=1,$$

and let $G_X(t) = E(t^X)$ be its PGF. An extra proportion of zeros $\rho \in (0,1)$ is added to the proportion of zeros from the distribution of the random variable $X$ while decreasing the remaining proportions in an appropriate way. The Zero-Inflated modification $Y$ of $X$ is defined by

$$P(Y=0) = \rho + (1-\rho) p_0$$
$$P(Y=j) = (1-\rho) p_j, \quad j=1,2,3,\ldots$$

It has as a PGF

$$G_Y(t) = \rho + (1-\rho)G_X(t)$$

In general, the inflation parameter $\rho$ may take negative values provided that

$$P(Y=0) > 0 \quad \text{i.e.,} \quad \rho \geq \frac{-p_0}{1-p_0} \quad \text{and therefore} \quad \max \left\{ \frac{-p_0}{1-p_0}, -1 \right\} \leq \rho \leq 0.$$

This case corresponds to the "opposite" phenomena - "excluding" a proportion of zeros from the basic discrete distribution, if necessary (see Kolev et al. 2000).

#### 5.3.1 Zero-Inflated Negative Binomial Distribution

The random variable $X$ defined by

$$P(X = 0) = \rho + (1-\rho) \left( \frac{a}{1+a} \right)^k$$
$$P(X = x) = (1-\rho) \left( \frac{\Gamma(k+x)a^k}{x!\Gamma(k)(1+a)^{k+x}} \right) \quad x=1,2,\ldots, a,k > 0$$
has Zero-Inflated Negative Binomial distribution with parameters $a, k > 0$ and

$$\rho \in \left\{ \max \left\{ \frac{-p_0}{1-p_0}, -1 \right\}, 1 \right\}.$$  

The mean, variance, and PGF of this distribution are given by

$$E(X) = \frac{k}{a} (1 - \rho)$$
$$Var(X) = \frac{k(1 - \rho)(1 + a - k\rho)}{a^2}$$
$$M_X(t) = \left( \frac{a}{1 + a - e^t} \right) (1 - \rho)$$

5.3.2 Zero-Inflated Discrete Weibull Distribution

The random variable $X$ defined by

$$P(X = 0) = \rho + (1 - \rho)(1 - e^{-\lambda})$$
$$P(X = x) = (1 - \rho) \left( e^{-\lambda x^m} - e^{-\lambda(x+1)^m} \right) \quad x = 1, 2, \ldots, \lambda, m > 0$$

has Zero-Inflated weibull distribution with parameters $\lambda, m > 0$,

$$\rho \in \left\{ \max \left\{ \frac{-p_0}{1-p_0}, -1 \right\}, 1 \right\}.$$  

6. DATA ANALYSIS

The data analyzed here are numbers of galaxies in a cubic window of the universe, provided to the authors by Prof. V.J. Martínez. The original window with side length of $254.014893 \times 133.134094 \times 31.072433 \, h^{-1} \times \text{Mpc}$ (1 Mpc is a million of parsecs, 1 parsec is 3.26 light years) has been divided into sub windows, cells, with side length of $4 \times 4 \times 4 \, h^{-1} \times \text{Mpc}$ and the numbers of galaxies were counted. Since the side length of the original window does not round number, the start point of division has been changed 5 times by a random shift. To combine these five data files in one, the rounded average frequency of these data sets was used in the analysis; see first two columns of Table 2. The empirical mean and variance of the cell counts are 0.526 and 1.41, respectively.

To study the dispersion of galaxies in the desired space and whether galaxies scattered through this space under the Poisson point pattern or not, it is used dispersion test introduced in section 3.1. The value of statistic is

$$I = 39253.001 > \chi^2_{(k-1, 0.95)} = 14833.743 \quad \text{so the hypothesis of Complete Spatial Randomness (CSR) is rejected at significant level } \alpha=0.05.$$  

The result shows that galaxies have been clustered in the space. The departure measurement of CRS is obtained by dispersion Indices provided in section 3.2, see Table 1.
Statistical analysis of the number of the galaxies in cubic cells in the universe

Table 1: Dispersion Indices

<table>
<thead>
<tr>
<th></th>
<th>ICS</th>
<th>ICF</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.697</td>
<td>1.697</td>
<td>0.31</td>
</tr>
</tbody>
</table>

In the following, the fitting results of all suggested distributions in section 4 are discussed in detail.

Table 2: Maximum likelihood estimates of the parameters of all suggested distributions for the galaxy data. The corresponding AICs of the models have been shown in the last column.

<table>
<thead>
<tr>
<th>Model</th>
<th>$\hat{\lambda}$ (SE)</th>
<th>$\hat{m}$ (SE)</th>
<th>$\hat{\mu}$ (SE)</th>
<th>$\hat{a}$ (SE)</th>
<th>$\hat{k}$ (SE)</th>
<th>AIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inf.Weibull</td>
<td>1.262198</td>
<td>1.411521</td>
<td>0.0546146</td>
<td>-</td>
<td>-</td>
<td>27450.62</td>
</tr>
<tr>
<td>Inf.Neg.bin</td>
<td>-</td>
<td>-</td>
<td>0.339(0.298)</td>
<td>0.543(0.041)</td>
<td>0.214(0.062)</td>
<td>27448.61</td>
</tr>
<tr>
<td>Neg.bin</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>0.605(0.021)</td>
<td>0.318(0.0095)</td>
<td>27449.19</td>
</tr>
<tr>
<td>Weibull</td>
<td>1.32(0.0136)</td>
<td>1.45(0.0181)</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>27448.94</td>
</tr>
</tbody>
</table>

Table 2 shows the estimated parameters of all studied models which were obtained by the maximum likelihood method using the Newton-Raphson optimization procedure, see Venables and Ripley (2002). The standard errors (SEs) given in parentheses were obtained via the diagonal of the inverse of the Hessian matrix of the log-likelihood function. About Zero-Inflated Weibull I was not able to obtain Hessian matrix. The SPLUS programs are available by request from the author.

Table 3: The empirical frequencies of numbers of the galaxies and the corresponding theoretical values of the studied distributions.

<table>
<thead>
<tr>
<th>Number of galaxies</th>
<th>Empirical numbers</th>
<th>Inf.weibull</th>
<th>Inf.Neg.bin</th>
<th>Neg.bin</th>
<th>Weibull</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>10659</td>
<td>10659.00</td>
<td>10659.00</td>
<td>10654.87</td>
<td>10666.85</td>
</tr>
<tr>
<td>1</td>
<td>2160</td>
<td>2144.84</td>
<td>2157.32</td>
<td>2159.69</td>
<td>2116.18</td>
</tr>
<tr>
<td>2</td>
<td>836</td>
<td>868.59</td>
<td>848.22</td>
<td>862.69</td>
<td>869.15</td>
</tr>
<tr>
<td>3</td>
<td>416</td>
<td>407.54</td>
<td>405.54</td>
<td>403.16</td>
<td>418.49</td>
</tr>
<tr>
<td>4</td>
<td>225</td>
<td>207.44</td>
<td>211.11</td>
<td>205.40</td>
<td>216.31</td>
</tr>
<tr>
<td>5</td>
<td>102</td>
<td>111.48</td>
<td>115.28</td>
<td>110.81</td>
<td>116.40</td>
</tr>
<tr>
<td>6</td>
<td>59</td>
<td>62.33</td>
<td>64.91</td>
<td>62.34</td>
<td>64.28</td>
</tr>
<tr>
<td>7</td>
<td>35</td>
<td>35.95</td>
<td>37.33</td>
<td>36.22</td>
<td>36.15</td>
</tr>
<tr>
<td>8</td>
<td>23</td>
<td>21.26</td>
<td>21.81</td>
<td>21.60</td>
<td>20.60</td>
</tr>
<tr>
<td>9</td>
<td>15</td>
<td>12.83</td>
<td>12.90</td>
<td>13.17</td>
<td>11.87</td>
</tr>
<tr>
<td>10</td>
<td>10</td>
<td>7.89</td>
<td>7.70</td>
<td>8.18</td>
<td>6.89</td>
</tr>
<tr>
<td>11</td>
<td>8</td>
<td>4.92</td>
<td>4.63</td>
<td>5.16</td>
<td>4.03</td>
</tr>
<tr>
<td>12</td>
<td>3</td>
<td>3.11</td>
<td>2.81</td>
<td>3.30</td>
<td>2.37</td>
</tr>
<tr>
<td>13</td>
<td>1</td>
<td>1.99</td>
<td>1.71</td>
<td>2.14</td>
<td>1.40</td>
</tr>
<tr>
<td>14</td>
<td>1</td>
<td>1.29</td>
<td>1.04</td>
<td>4.26</td>
<td>2.05</td>
</tr>
</tbody>
</table>
The difference between empirical frequencies and theoretical values of the fitted distributions.

To compare the goodness-of-fit of the two models the Akaike Information Criterion (AIC) is used. The AIC is a likelihood-based statistic and includes a penalty for each parameter in the model (Akaike (1983)) and is computed as twice the negative likelihood plus twice the number of estimated parameters; a lower AIC indicates a better fit. Ghorbani et al. (2006) showed that Weibull distribution fits to the number of galaxies in cubic cells better than Negative Binomial. In Table 1, the Comparison of the AICs of four studied models shows that the Zero-Inflated NB fits the data better than the discrete Weibull distribution. The empirical and theoretical numbers are shown in Table 3.

7. COMMENTS AND CONCLUSION

AIC criteria, Table 3 and Figure 1 show that Zero-Inflated NB distribution fits data better than other suggested models. Moreover, this distribution properly estimates the number of cells which has no galaxies in. Thus, according to the importance of VPF estimation in cosmology, this property can be considered one of the strengths of this distribution.

Serfozo (1990) introduced two point processes in which the number of points in disjoint cells has a negative binomial distribution, see also Diggle and Milne (1983). But probably until now no point processes have been introduced in which the number of points in disjoint cells has a discrete Weibull distribution and Zero-Inflated distributions. The question remains open to interpret the parameters $\lambda$ and $m$ and $\rho$ cosmologically.
REFERENCES

ABSTRACT

Neighbor balanced designs ensure that treatment comparisons will be as less affected by neighbor effects as possible. In field experiments, smaller blocks have more practical applications. This article deals with the construction of universally optimal neighbor designs in circular blocks of size three. Catalogues of the proposed designs are also compiled.

MSC(2000): 05B05, 62K10

KEYWORDS

Circular binary blocks; Generalized neighbor designs; Neighbor balanced designs; Universally optimal neighbor designs.

1. INTRODUCTION

The design strategy of a statistical experiment is influenced to a large extent by the nature of dependence that exists among the observations. Experiments in agriculture, horticulture and forestry often show neighbor effects (Azais et al., 1993). Rees (1967) introduced neighbor designs in serology. Rees (1967), Hwang (1973), Lawless (1971), Das and Saha (1976), Bermond and Faber (1976), Dey and Chakravarty (1977), Azais et al. (1993), Ahmed and Akhtar (2008), and Iqbal et al. (2009) constructed neighbor designs for several configurations. Iqbal et al. (2006) constructed several second order neighbor designs using cyclic shifts. Ai et al. (2007) constructed all order neighbor balanced designs for \( v \) prime or prime power when \( k \leq v \). Optimality of circular neighbor balanced designs for total effects with autoregressive correlated observations is discussed by Ai et al. (2009). Akhtar and Ahmed (2009) proposed new second and higher order neighbor designs in circular binary blocks. Ahmed et al. (2009) generated economical generalized neighbor designs of use in serology.

Neighbor designs for all \( v \) in linear blocks of size 3 are constructed by Jacroux (1998). Akhtar et al. (2010) presented a catalogue of neighbor designs in circular blocks of size five. Ahmed and Akhtar (2011) constructed neighbor designs in circular blocks of size six. In this article, neighbor designs are constructed in circular blocks of size three.

Kiefer (1975) introduced the universal optimality criterion which includes the well known D, A and E optimality criteria as special case. Kunert (1984) showed that circular
neighbor designs in $\Omega_{(v,v-1,v)}$ are universally optimal for the estimation of treatments as well as neighbor effects. Druilhet (1999) generalized the result of Kunert (1984) such that a circular neighbor balanced design (CNBD) is universally optimal for the estimation of treatment effects as well as neighbor effects over all designs in $\Omega_{(v,b,k)}$ for $3 \leq k \leq v$.

According to Druilhet (1999), a design which satisfies the following conditions is CNBD.

i) A circular design in which each treatment is equireplicated and appears at most once in each block.

ii) A balanced block design.

iii) Each ordered pair of distinct treatments appears equal number of times as neighbors.

Bailey and Druilhet (2004), Filipiak and Markiewicz (2005) and Filipiak and Rozanski (2005) also showed that circular neighbor balanced designs are universally optimal. Ai et al. (2007) showed that a circular block design neighbor balanced at distances up to $\gamma \leq k-1$ is universally optimal for total effects.

Our proposed designs satisfy all the conditions of CNBD, therefore, these designs are universally optimal and are constructed under the mathematical model:

$$Y = X_0\mu + X_1\tau + X_2\beta + \varepsilon$$

where $Y$ is the bk x 1 vector of response.

$X_0$ is the bk x 1 vector of 1’s.

$X_1$ is the bk x v incidence matrix for treatment effects.

$X_2$ is the bk x v incidence matrix for first-order neighbor effects.

$\tau$ is the v x 1 vector of treatment effects.

$\beta$ is the v x 1 vector of first-order neighbor effects.

$\varepsilon$ is the bk x 1 vector of random errors.

- If all off-diagonals values of $X_2'X_2$ are same, say $\lambda'$ times then design is called neighbor balanced. If $\lambda' = 1$, design is minimal neighbor balanced.

- If all off-diagonals values of $X_2'X_2$ are not same, design is called generalized neighbor design. If all off-diagonals values of $X_2'X_2$ are either $\lambda'_1$ or $\lambda'_2$ with $\lambda'_2 = \lambda'_1 + 1$, design is called GN$2$-designs. GN$2$-designs are thought to be closest to the neighbor designs. If $\lambda'_1 = 1$ and $\lambda'_2 = 2$, GN$2$-designs are minimal. In Section 3, minimal GN$2$-designs are constructed for the cases where neighbor designs required a large number of blocks.

**2. NEIGHBOR BALANCED DESIGNS FOR $v$ ODD AND $k = 3$**

In this section, neighbor balanced designs (NBD) are constructed in circular blocks of size 3 for $v$ odd.
Series 2.1

NBD for \( v = 2m+1 \) can be constructed in \( mv \) circular blocks of size 3 by developing the following \( m \) initial blocks cyclically mod \( v \).

\[ I_j = (0, j, 2j); \quad j = 1, 2, \ldots, m. \]

**Proof.**
The combined set of forward and backward differences between neighboring elements take all the values from 1 to \( v-1 \) equally often. Hence the theorem.

**Example 2.1**

NBD is generated for \( v = 17 \) and \( k = 3 \) by developing the following eight initial blocks cyclically mod 17.

\[ I_1 = (0,1,2), \quad I_2 = (0,2,4), \quad I_3 = (0,3,6), \quad I_4 = (0,4,8) \]
\[ I_5 = (0,5,10), \quad I_6 = (0,6,12), \quad I_7 = (0,7,14), \quad I_8 = (0,8,16) \]

These designs can also be constructed in less than \( mv \) blocks if \( v \) or \((v-1)\) is divisible by 3.

Series 2.2

Minimal NBD for \( v = 6t+1 \), \( t \) integer can be constructed in \( tv \) circular blocks of size 3 by developing \( t \) initial blocks cyclically mod \( v \).

**Example 2.2**

NBD is generated for \( v = 19 \) and \( k = 3 \) by developing the following three initial blocks cyclically mod 19.

\[ I_1 = (0,1,5), \quad I_2 = (0,2,8), \quad I_3 = (0,3,10) \]

Series 2.3.

Minimal NBD for \( v = 3t; \ t(>1) \) odd can be constructed in \( tv \) circular blocks of size 3 by developing \((t-1)/2 \) initial blocks cyclically mod \( v \) along with \( v/3 \) augmented blocks.

**Example 2.3**

NBD is generated for \( v = 15 \) and \( k = 3 \) by developing the following two initial blocks cyclically mod 15 along with five augmented blocks.

\[ I_1 = (0,1,4), \quad I_2 = (0,2,8) \quad \text{Augmented blocks} \]
\[ (0,5,10), (1,6,11), (2,7,12), (3,8,13), (4,9,14) \]

3. NEIGHBOR BALANCED DESIGNS FOR \( v \) EVEN AND \( k = 3 \)

Series 3.1

NBD for \( v = 2m+2 \) can be constructed in \( v(v-1) \) circular blocks of size 3 by developing the following \( v \) initial blocks cyclically mod \((v-1)\).

\[ I_j = (0, j, 2j); \quad j = 1, 2, \ldots, m-1. \]
\[ I_{j+m-1} = I_j; \quad j = 1, 2, \ldots, m-1. \]
\[ I_{2m-1} = (0, m, 2m), \]
\[ I_{2m+1} = I_{2m} = (0, m, \infty), \]
\[ I_{2m+2} = (0, 1, \infty), \quad \text{where} \ \infty = v-1. \]

These designs can also be constructed in less than \( v(v-1) \) blocks if \( v \) or \((v-1)\) is divisible by 3.
Case (i). NBD when $(v-1)$ is odd and divisible by 3

NBD can be constructed for $v = 3i+1$; $i$ odd and $k = 3$ in $iv$ circular blocks by developing $i$ initial blocks cyclically mod $v$.

Example 2.4

NBD is generated for $v = 16$ and $k = 3$ by developing the following five initial blocks cyclically mod $16$.

$I_1 = I_2 = (0,1,7)$, $I_3 = I_4 = (0,5,3)$, $I_5 = (0,8,12)$

Case (ii). NBD when $v$ is even and divisible by 3

NBD can be constructed for $v = 3i$; $i$ even and $k = 3$ by developing $i$ initial blocks (one of these blocks contains $\infty$) cyclically mod $(v-1)$.

Example 2.5

NBD is generated for $v = 18$ and $k = 3$ by developing the following six initial blocks cyclically mod $17$.

$I_1 = I_2 = (0,1,3)$, $I_3 = I_4 = (0,4,10)$,
$I_5 = (0,5,11)$, $I_6 = (0,5,\infty)$, where $\infty = 17$.

Following is the catalogue of neighbor designs when $v$ or $(v-1)$ is divisible by $k = 3$.

<table>
<thead>
<tr>
<th>$v$</th>
<th>Initial Block</th>
</tr>
</thead>
<tbody>
<tr>
<td>18</td>
<td>$(0,1,4),(0,2,7),(0,2,6),(0,1,7),(0,3,8),(0,8,\infty)$</td>
</tr>
<tr>
<td>22</td>
<td>$(0,1,10),(0,2,9),(0,3,8),(0,3,8),(0,2,6),(0,4,10),(0,1,\infty)$</td>
</tr>
<tr>
<td></td>
<td><strong>Augmented blocks</strong></td>
</tr>
<tr>
<td></td>
<td>$(0,7,14),(1,8,15),(2,9,16),(3,10,17),(4,11,18),(5,12,19),(6,13,20)$</td>
</tr>
<tr>
<td>24</td>
<td>$(0,1,11),(0,2,9),(0,2,9),(0,3,8),(0,4,10),(0,5,11),(0,1,4),(0,8,\infty)$</td>
</tr>
<tr>
<td>28</td>
<td>$(0,1,14),(0,2,12)\infty,(0,3,11)\infty,(0,4,9)\infty,(0,6,13),(0,1,7)$</td>
</tr>
<tr>
<td>30</td>
<td>$(0,1,14),(0,2,12)\infty,(0,3,11)\infty,(0,4,9)\infty,(0,6,13),(0,1,7),(0,14,\infty)$</td>
</tr>
<tr>
<td>33</td>
<td>$(0,2,12),(0,4,13),(0,7,15),(0,1,6),(0,3,17)$</td>
</tr>
<tr>
<td></td>
<td><strong>Augmented blocks</strong></td>
</tr>
<tr>
<td></td>
<td>$(0,11,22),(1,12,23),(2,13,24),(3,14,25),(4,15,26),(5,16,27),(6,17,28),(7,18,29)$</td>
</tr>
<tr>
<td></td>
<td>$(8,19,30),(9,20,31),(10,21,32)$</td>
</tr>
<tr>
<td>34</td>
<td>$(0,2,12),(0,4,13),(0,7,15),(0,1,6),(0,3,17),(0,2,12),(0,4,13),(0,7,15),(0,1,6)$</td>
</tr>
<tr>
<td></td>
<td><strong>Augmented blocks</strong></td>
</tr>
<tr>
<td></td>
<td>$(0,11,22),(1,12,23),(2,13,24),(3,14,25),(4,15,26),(5,16,27),(6,17,28),(7,18,29)$</td>
</tr>
<tr>
<td></td>
<td>$(8,19,30),(9,20,31),(10,21,32)$</td>
</tr>
<tr>
<td>36</td>
<td>$(0,1,17),(0,2,15),(0,3,12),(0,4,11),(0,5,10),(0,6,14),(0,1,17),(0,2,15),(0,3,12)$</td>
</tr>
<tr>
<td></td>
<td><strong>Augmented blocks</strong></td>
</tr>
<tr>
<td></td>
<td>$(0,4,11),(0,5,10),(0,6,14),(0,10,\infty)$</td>
</tr>
<tr>
<td>39</td>
<td>$(0,2,10),(0,3,14),(0,5,12),(0,6,15),(0,1,18),(0,4,20)$</td>
</tr>
<tr>
<td></td>
<td><strong>Augmented blocks</strong></td>
</tr>
<tr>
<td></td>
<td>$(0,13,26),(1,14,27),(2,15,28),(3,16,29),(4,17,30),(5,18,31),(6,19,32),(7,20,33)$</td>
</tr>
<tr>
<td></td>
<td>$(8,21,34),(9,22,35),(10,23,36),(11,24,37),(12,25,38)$</td>
</tr>
<tr>
<td>( v )</td>
<td>Initial Block</td>
</tr>
<tr>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>40</td>
<td>((0,2,10),(0,3,14),(0,5,12),(0,6,15),(0,1,18),(0,4,20),(0,2,10),(0,3,14),(0,5,12),(0,6,15),(0,1,18),(0,4,20),(0,13,\infty))</td>
</tr>
<tr>
<td>42</td>
<td>((0,1,20),(0,1,20),(0,2,18),(0,3,15),(0,4,14),(0,5,13),(0,6,17),(0,6,17),(0,16,34),(0,2,9),(0,9,\infty))</td>
</tr>
<tr>
<td>43</td>
<td>((0,2,20),(0,3,22),(0,4,17),(0,5,15),(0,6,14),(0,1,12),(0,7,16))</td>
</tr>
<tr>
<td>45</td>
<td>((0,2,21),(0,3,20),(0,4,18),(0,5,16),(0,6,13),(0,1,9),(0,10,22))</td>
</tr>
<tr>
<td>46</td>
<td>((0,1,23),(0,7,19),(0,3,20),(0,4,18),(0,5,16),(0,2,10),(0,12,25))</td>
</tr>
<tr>
<td>48</td>
<td>((0,1,24),(0,2,22),(0,3,18),(0,4,11),(0,5,21),(0,6,14),(0,9,19),(0,12,25))</td>
</tr>
<tr>
<td>49</td>
<td>((0,3,23),(0,2,21),(0,10,27),(0,4,18),(0,5,16),(0,1,8),(0,6,15),(0,12,25))</td>
</tr>
<tr>
<td>51</td>
<td>((0,1,24),(0,2,22),(0,3,18),(0,4,11),(0,5,21),(0,6,14),(0,9,19),(0,12,25))</td>
</tr>
<tr>
<td>52</td>
<td>((0,1,24),(0,2,22),(0,3,18),(0,4,11),(0,5,21),(0,6,14),(0,9,19),(0,12,25),(0,17,\infty))</td>
</tr>
<tr>
<td>54</td>
<td>((0,1,16),(0,2,22),(0,3,8),(0,4,28),(0,7,18),(0,9,23),(0,10,27),(0,13,32),(0,16,\infty))</td>
</tr>
<tr>
<td>55</td>
<td>((0,1,26),(0,2,23),(0,4,24),(0,5,19),(0,6,17),(0,7,15),(0,3,16),(0,9,27),(0,10,22))</td>
</tr>
<tr>
<td>57</td>
<td>((0,2,14),(0,3,26),(0,4,24),(0,5,27),(0,6,21),(0,7,25),(0,8,17),(0,11,\infty),(0,13,29))</td>
</tr>
<tr>
<td>58</td>
<td>((0,2,14),(0,3,26),(0,4,24),(0,5,27),(0,6,21),(0,7,25),(0,8,17),(0,11,\infty),(0,13,29),(0,19,\infty))</td>
</tr>
</tbody>
</table>
 Universally optimal neighbor designs in blocks of size three

<table>
<thead>
<tr>
<th>v</th>
<th>Initial Block</th>
<th>Augmented blocks</th>
</tr>
</thead>
<tbody>
<tr>
<td>60</td>
<td>(0,1,28),(0,3,26),(0,4,25),(0,6,24),(0,8,22),(0,9,20),(0,7,19),(0,13,29),(0,2,17), (0,5,10),(0,10,∞)</td>
<td>(0,21,42),(1,22,43),(2,23,44),(3,24,45),(4,25,46),(5,26,47),(6,27,48),(7,28,49), (8,29,50),(9,30,51),(10,31,52),(11,32,53),(12,33,54),(13,34,55),(14,35,56), (15,36,57),(16,37,58),(17,38,59),(18,39,60),(19,40,61),(20,41,62)</td>
</tr>
<tr>
<td>61</td>
<td>(0,1,29),(0,25,52),(0,3,26),(0,4,24),(0,5,22),(0,6,21),(0,7,19),(0,8,18),(0,2,13), (0,14,30)</td>
<td>(0,21,42),(1,22,43),(2,23,44),(3,24,45),(4,25,46),(5,26,47),(6,27,48),(7,28,49), (8,29,50),(9,30,51),(10,31,52),(11,32,53),(12,33,54),(13,34,55),(14,35,56), (15,36,57),(16,37,58),(17,38,59),(18,39,60),(19,40,61),(20,41,62)</td>
</tr>
<tr>
<td>63</td>
<td>(0,1,8),(0,3,9),(0,11,30),(0,12,29),(0,14,41),(0,15,38),(0,17,45),(0,24,4),(0,26,5)</td>
<td>(0,21,42),(1,22,43),(2,23,44),(3,24,45),(4,25,46),(5,26,47),(6,27,48),(7,28,49), (8,29,50),(9,30,51),(10,31,52),(11,32,53),(12,33,54),(13,34,55),(14,35,56), (15,36,57),(16,37,58),(17,38,59),(18,39,60),(19,40,61),(20,41,62)</td>
</tr>
<tr>
<td>64</td>
<td>(0,1,8),(0,3,9),(0,11,30),(0,12,29),(0,14,41),(0,15,38),(0,17,45),(0,24,4),(0,26,5)</td>
<td>(0,21,42),(1,22,43),(2,23,44),(3,24,45),(4,25,46),(5,26,47),(6,27,48),(7,28,49), (8,29,50),(9,30,51),(10,31,52),(11,32,53),(12,33,54),(13,34,55),(14,35,56), (15,36,57),(16,37,58),(17,38,59),(18,39,60),(19,40,61),(20,41,62)</td>
</tr>
<tr>
<td>66</td>
<td>(0,1,31),(0,2,29),(0,4,28),(0,5,26),(0,7,25),(0,8,23),(0,10,22),(0,9,20),(0,3,17),(0,6,19),(0,1,31),(0,2,29),(0,4,28),(0,5,26),(0,7,25),(0,8,23),(0,10,22),(0,9,20),(0,3,17),(0,6,19)</td>
<td>(0,21,42),(1,22,43),(2,23,44),(3,24,45),(4,25,46),(5,26,47),(6,27,48),(7,28,49), (8,29,50),(9,30,51),(10,31,52),(11,32,53),(12,33,54),(13,34,55),(14,35,56), (15,36,57),(16,37,58),(17,38,59),(18,39,60),(19,40,61),(20,41,62)</td>
</tr>
<tr>
<td>67</td>
<td>(0,1,32),(0,2,30),(0,4,29),(0,5,26),(0,6,24),(0,7,15),(0,9,22),(0,3,19),(0,10,27),(0,11,23),(0,14,34)</td>
<td>(0,21,42),(1,22,43),(2,23,44),(3,24,45),(4,25,46),(5,26,47),(6,27,48),(7,28,49), (8,29,50),(9,30,51),(10,31,52),(11,32,53),(12,33,54),(13,34,55),(14,35,56), (15,36,57),(16,37,58),(17,38,59),(18,39,60),(19,40,61),(20,41,62)</td>
</tr>
<tr>
<td>69</td>
<td>(0,1,7),(0,3,8),(0,11,35),(0,12,33),(0,13,17),(0,14,31),(0,30,50),(0,29,47),(0,28,2),(0,16,60),(0,27,59)</td>
<td>(0,23,46),(1,24,47),(2,25,48),(3,26,49),(4,27,50),(5,28,51),(6,29,52),(7,30,53),(8,31,54),(9,32,55),(10,33,56),(11,34,57),(12,35,58),(13,36,59),(14,37,60),(15,38,61),(16,39,62),(17,40,63),(18,41,64),(19,42,65),(20,43,66),(21,44,67),(22,45,68)</td>
</tr>
<tr>
<td>70</td>
<td>(0,1,7),(0,3,8),(0,11,35),(0,12,33),(0,13,17),(0,14,31),(0,30,50),(0,29,47),(0,28,2),(0,16,60),(0,27,59),(0,23,∞)</td>
<td>(0,23,46),(1,24,47),(2,25,48),(3,26,49),(4,27,50),(5,28,51),(6,29,52),(7,30,53),(8,31,54),(9,32,55),(10,33,56),(11,34,57),(12,35,58),(13,36,59),(14,37,60),(15,38,61),(16,39,62),(17,40,63),(18,41,64),(19,42,65),(20,43,66),(21,44,67),(22,45,68)</td>
</tr>
<tr>
<td>72</td>
<td>(0,1,34),(0,2,31),(0,5,30),(0,7,27),(0,6,15),(0,10,23),(0,12,26),(0,11,32),(0,17,35),(0,3,22),(0,4,28),(0,8,16),(0,1,34),(0,2,31),(0,5,30),(0,7,27),(0,6,15),(0,10,23),(0,12,26),(0,11,32),(0,17,35),(0,3,22),(0,4,28),(0,8,16),(0,16,∞)</td>
<td>(0,23,46),(1,24,47),(2,25,48),(3,26,49),(4,27,50),(5,28,51),(6,29,52),(7,30,53),(8,31,54),(9,32,55),(10,33,56),(11,34,57),(12,35,58),(13,36,59),(14,37,60),(15,38,61),(16,39,62),(17,40,63),(18,41,64),(19,42,65),(20,43,66),(21,44,67),(22,45,68)</td>
</tr>
<tr>
<td>73</td>
<td>(0,1,35),(0,2,33),(0,3,30),(0,4,8),(0,6,24),(0,7,23),(0,8,20),(0,10,21),(0,13,41),(0,14,29),(0,17,36),(0,22,47)</td>
<td>(0,23,46),(1,24,47),(2,25,48),(3,26,49),(4,27,50),(5,28,51),(6,29,52),(7,30,53),(8,31,54),(9,32,55),(10,33,56),(11,34,57),(12,35,58),(13,36,59),(14,37,60),(15,38,61),(16,39,62),(17,40,63),(18,41,64),(19,42,65),(20,43,66),(21,44,67),(22,45,68)</td>
</tr>
<tr>
<td>$v$</td>
<td>Initial Block</td>
<td></td>
</tr>
<tr>
<td>-----</td>
<td>---------------</td>
<td></td>
</tr>
<tr>
<td>75</td>
<td>(0,12,38),(0,11,35),(0,13,36),(0,14,34),(0,15,33),(0,32,59),(0,31,9),(0,30,2), (0,29,10),(0,21,4), (0,1,7),(0,3,8)</td>
<td></td>
</tr>
<tr>
<td></td>
<td><strong>Augmented blocks</strong></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0,25,50),(1,26,51),(2,27,52),(3,28,53),(4,29,54),(5,30,55),(6,31,56),(7,32,57), (8,33,58),(9,34,59),(10,35,60),(11,36,61),(12,37,62),(13,38,63),(14,39,64), (15,40,65),(16,41,66),(17,42,67),(18,43,68),(19,44,69),(20,45,70),(21,46,71), (22,47,72),(23,48,73),(24,49,74)</td>
<td></td>
</tr>
<tr>
<td>76</td>
<td>(0,12,38),(0,11,35),(0,13,36),(0,14,34),(0,15,33),(0,32,59),(0,31,9),(0,30,2), (0,29,10),(0,21,4),(0,1,7),(0,3,8),(0,12,38),(0,11,35),(0,13,36),(0,14,34),(0,15,33),(0,32,59),(0,31,9),(0,30,2), (0,29,10),(0,21,4),(0,1,7),(0,3,8),(0,25,∞)</td>
<td></td>
</tr>
<tr>
<td></td>
<td><strong>Augmented blocks</strong></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0,25,50),(1,26,51),(2,27,52),(3,28,53),(4,29,54),(5,30,55),(6,31,56),(7,32,57), (8,33,58),(9,34,59),(10,35,60),(11,36,61),(12,37,62),(13,38,63),(14,39,64), (15,40,65),(16,41,66),(17,42,67),(18,43,68),(19,44,69),(20,45,70),(21,46,71), (22,47,72),(23,48,73),(24,49,74)</td>
<td></td>
</tr>
<tr>
<td>78</td>
<td>(0,1,37),(0,2,35),(0,4,34),(0,5,32),(0,7,31),(0,8,29),(0,10,28),(0,11,26),(0,12,25),(0,3,20),(0,6,22),(0,9,23),(0,1,37),(0,2,35),(0,4,34),(0,5,32),(0,7,31),(0,8,29),(0,10,28),(0,11,26),(0,12,25),(0,3,20),(0,6,22),(0,9,23),(0,19,38),(0,38,∞)</td>
<td></td>
</tr>
<tr>
<td>79</td>
<td>(0,1,14),(0,2,36),(0,3,35),(0,4,41),(0,5,31),(0,6,30),(0,7,29),(0,8,28),(0,9,25),(0,10,27),(0,11,23),(0,15,33),(0,19,40)</td>
<td></td>
</tr>
<tr>
<td>81</td>
<td>(0,12,41),(0,11,39),(0,13,38),(0,14,37),(0,15,36),(0,16,35),(0,34,64),(0,33,59), (0,32,63),(0,24,44),(0,18,28),(0,2,7),(0,3,9)</td>
<td></td>
</tr>
<tr>
<td></td>
<td><strong>Augmented blocks</strong></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0,27,54),(1,28,55),(2,29,56),(3,30,57),(4,31,58),(5,32,59),(6,33,60),(7,34,61), (8,35,62),(9,36,63),(10,37,64),(11,38,65),(12,39,66),(13,40,67),(14,41,68), (15,42,69),(16,43,70),(17,44,71),(18,45,72),(19,46,73),(20,47,74),(21,48,75), (22,49,76),(23,50,77),(24,51,78),(25,52,79),(26,53,80)</td>
<td></td>
</tr>
<tr>
<td>82</td>
<td>(0,12,41),(0,11,39),(0,13,38),(0,14,37),(0,15,36),(0,16,35),(0,34,64),(0,33,59), (0,32,63),(0,24,44),(0,18,28),(0,2,7),(0,3,9),(0,12,41),(0,11,39),(0,13,38),(0,14,37),(0,15,36),(0,16,35),(0,34,64),(0,33,59),(0,32,63),(0,24,44),(0,18,28),(0,2,7),(0,3,9),(0,27,∞)</td>
<td></td>
</tr>
<tr>
<td></td>
<td><strong>Augmented blocks</strong></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0,27,54),(1,28,55),(2,29,56),(3,30,57),(4,31,58),(5,32,59),(6,33,60),(7,34,61), (8,35,62),(9,36,63),(10,37,64),(11,38,65),(12,39,66),(13,40,67),(14,41,68), (15,42,69),(16,43,70),(17,44,71),(18,45,72),(19,46,73),(20,47,74),(21,48,75), (22,49,76),(23,50,77),(24,51,78),(25,52,79),(26,53,80)</td>
<td></td>
</tr>
<tr>
<td>84</td>
<td>(0,1,40),(0,2,37),(0,3,22),(0,4,18),(0,5,34),(0,6,31),(0,7,30),(0,8,28),(0,9,42), (0,10,27),(0,11,24),(0,15,36),(0,12,38),(0,1,40),(0,2,37),(0,3,22),(0,4,18),(0,5,34),(0,6,31),(0,7,30),(0,8,28),(0,9,42),(0,10,27),(0,11,24),(0,15,36),(0,12,38),(0,16,32),(0,3,2∞)</td>
<td></td>
</tr>
<tr>
<td>85</td>
<td>(0,1,13),(0,2,39),(0,3,38),(0,4,44),(0,5,34),(0,6,27),(0,7,32),(0,8,31),(0,9,28),(0,10,30),(0,11,26),(0,14,36),(0,16,33),(0,18,42)</td>
<td></td>
</tr>
<tr>
<td>$v$</td>
<td>Initial Block</td>
<td></td>
</tr>
<tr>
<td>-----</td>
<td>---------------</td>
<td></td>
</tr>
<tr>
<td>87</td>
<td>(0,2,39),(0,4,42),(0,6,36),(0,5,32),(0,7,28),(0,9,31),(0,10,26),(0,11,25),(0,1,24), (0,15,34),(0,17,35),(0,3,43),(0,8,20),(0,13,46)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Augmented blocks</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0,29,58),(1,30,59),(2,31,60),(3,32,61),(4,33,62),(5,34,63),(6,35,64),(7,36,65), (8,37,66),(9,38,67),(10,39,68),(11,40,69),(12,41,70),(13,42,71),(14,43,72), (15,44,73),(16,45,74),(17,46,75),(18,47,76),(19,48,77),(20,49,78),(21,50,79), (22,51,80),(23,52,81),(24,53,82),(25,54,83),(26,55,84),(27,56,85),(28,57,86)</td>
<td></td>
</tr>
<tr>
<td>88</td>
<td>(0,2,39),(0,4,42),(0,6,36),(0,5,32),(0,7,28),(0,9,31),(0,10,26),(0,11,25),(0,1,24), (0,15,34),(0,17,35),(0,3,43),(0,8,20),(0,13,46)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Augmented blocks</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0,29,58),(1,30,59),(2,31,60),(3,32,61),(4,33,62),(5,34,63),(6,35,64),(7,36,65), (8,37,66),(9,38,67),(10,39,68),(11,40,69),(12,41,70),(13,42,71),(14,43,72), (15,44,73),(16,45,74),(17,46,75),(18,47,76),(19,48,77),(20,49,78),(21,50,79), (22,51,80),(23,52,81),(24,53,82),(25,54,83),(26,55,84),(27,56,85),(28,57,86)</td>
<td></td>
</tr>
<tr>
<td>90</td>
<td>(0,1,13),(0,2,41),(0,3,25),(0,4,47),(0,5,36),(0,6,35),(0,7,33),(0,8,32),(0,9,23), (0,10,30),(0,11,28),(0,16,37),(0,18,45)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Augmented blocks</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0,1,13),(0,2,41),(0,3,25),(0,4,47),(0,5,36),(0,6,35),(0,7,33),(0,8,32),(0,9,23), (0,10,30),(0,11,28),(0,16,37),(0,18,45)</td>
<td></td>
</tr>
<tr>
<td>91</td>
<td>(0,1,14),(0,2,42),(0,3,41),(0,4,19),(0,5,37),(0,6,36),(0,7,35),(0,8,34),(0,9,25), (0,10,33),(0,11,21),(0,12,29),(0,13,27),(0,18,39),(0,22,46)</td>
<td></td>
</tr>
<tr>
<td>93</td>
<td>(0,12,47),(0,11,45),(0,13,43),(0,15,44),(0,14,42),(0,16,41),(0,17,40),(0,18,39), (0,36,73),(0,5,38),(0,6,32),(0,1,10),(0,8,27),(0,2,24),(0,3,7)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Augmented blocks</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0,12,47),(0,11,45),(0,13,43),(0,15,44),(0,14,42),(0,16,41),(0,17,40),(0,18,39), (0,36,73),(0,5,38),(0,6,32),(0,1,10),(0,8,27),(0,2,24),(0,3,7)</td>
<td></td>
</tr>
<tr>
<td>94</td>
<td>(0,12,47),(0,11,45),(0,13,43),(0,15,44),(0,14,42),(0,16,41),(0,17,40),(0,18,39), (0,36,73),(0,5,38),(0,6,32),(0,1,10),(0,8,27),(0,2,24),(0,3,7)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Augmented blocks</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0,12,47),(0,11,45),(0,13,43),(0,15,44),(0,14,42),(0,16,41),(0,17,40),(0,18,39), (0,36,73),(0,5,38),(0,6,32),(0,1,10),(0,8,27),(0,2,24),(0,3,7)</td>
<td></td>
</tr>
<tr>
<td>96</td>
<td>(0,1,17),(0,2,44),(0,3,40),(0,4,49),(0,5,39),(0,6,32),(0,7,38),(0,8,27),(0,9,22), (0,10,30),(0,11,23),(0,14,35),(0,15,48),(0,24,52),(0,25,54),(0,1,17),(0,2,44), (0,3,40),(0,4,49),(0,5,39),(0,6,32),(0,7,38),(0,8,27),(0,9,22),(0,10,30),(0,11,23), (0,14,35),(0,15,48),(0,24,52),(0,25,54),(0,18,36),(0,36,∞)</td>
<td></td>
</tr>
<tr>
<td>97</td>
<td>(0,1,17),(0,2,45),(0,3,44),(0,4,50),(0,5,40),(0,6,20),(0,7,38),(0,8,37),(0,9,36), (0,10,34),(0,11,32),(0,12,30),(0,13,28),(0,19,42),(0,22,48),(0,33,72)</td>
<td></td>
</tr>
</tbody>
</table>
In this section, generalized neighbor designs are constructed in circular blocks of size three with \( \lambda_1' = 1 \) and \( \lambda_2' = 2 \), called GN_2-designs.

**Case (i). GN_2-designs for \( v \) odd and neither \( v \) nor \( (v-1) \) divisible by 3**

GN_2-designs can be constructed for \( v = 3i+2; \) \( i \) even and \( k = 3 \) in circular blocks by developing \( (i+2)/2 \) initial blocks cyclically mod \( v \).

**Example 3.1**

GN_2-designs is generated for \( v = 32 \) and \( k = 3 \) by developing the following five initial blocks cyclically mod 32.

\[
I_1 = (0,1,3), \quad I_2 = (0,4,9), \quad I_3 = (0,6,16), \\
I_4 = (0,7,15), \quad I_5 = (0,11,23), \quad I_6 = (0,13,27)
\]

**Case (ii). GN_2-designs for \( v \) even and neither \( v \) nor \( (v-1) \) divisible by 3**

GN_2-designs is constructed for \( v = 3i+2; \) \( i \) odd and \( k = 3 \) in circular blocks by developing \( i \) initial blocks cyclically mod \( v \).

**Example 3.2**

GN_2-designs is generated for \( v = 35 \) and \( k = 3 \) by developing the following six initial blocks cyclically mod 17.

\[
I_1 = (0,1,3), \quad I_2 = (0,2,14), \quad I_3 = (0,3,14), \\
I_4 = (0,4,9), \quad I_5 = (0,6,13), \quad I_6 = (0,8,18)
\]

Following is the catalogue of GN_2-designs for \( k = 3 \).
<table>
<thead>
<tr>
<th>$V$</th>
<th>Initial Block(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>(0,1,3)</td>
</tr>
<tr>
<td>6</td>
<td>(0,1,3)</td>
</tr>
<tr>
<td>8</td>
<td>(0,1,3), (0,3,7)</td>
</tr>
<tr>
<td>11</td>
<td>(0,1,3), (0,4,9)</td>
</tr>
<tr>
<td>14</td>
<td>(0,1,3), (0,4,9), (0,6,13)</td>
</tr>
<tr>
<td>17</td>
<td>(0,1,3), (0,4,9), (0,6,13)</td>
</tr>
<tr>
<td>20</td>
<td>(0,1,3), (0,4,9), (0,6,13), (0,8,18)</td>
</tr>
<tr>
<td>22</td>
<td>(0,1,3), (0,4,12), (0,5,11), (0,7,16)</td>
</tr>
<tr>
<td>23</td>
<td>(0,1,3), (0,4,9), (0,6,13), (0,8,19)</td>
</tr>
<tr>
<td>24</td>
<td>(0,1,3), (0,4,12), (0,6,13), (0,5,14)</td>
</tr>
<tr>
<td>26</td>
<td>(0,1,3), (0,4,9), (0,6,13), (0,8,18), (0,11,23)</td>
</tr>
<tr>
<td>28</td>
<td>(0,1,3), (0,4,14), (0,5,11), (0,7,15), (0,9,21)</td>
</tr>
<tr>
<td>29</td>
<td>(0,1,3), (0,4,9), (0,6,13), (0,8,18), (0,12,26)</td>
</tr>
<tr>
<td>30</td>
<td>(0,1,3), (0,4,12), (0,6,13), (0,4,15), (0,8,20)</td>
</tr>
<tr>
<td>32</td>
<td>(0,1,3), (0,4,9), (0,6,16), (0,7,15), (0,11,23), (0,13,27)</td>
</tr>
<tr>
<td>34</td>
<td>(0,1,3), (0,4,9), (0,6,17), (0,7,19), (0,8,18), (0,13,17)</td>
</tr>
<tr>
<td>35</td>
<td>(0,1,3), (0,4,9), (0,6,13), (0,8,18), (0,2,14), (0,3,14)</td>
</tr>
<tr>
<td>38</td>
<td>(0,1,3), (0,4,9), (0,6,19), (0,7,15), (0,10,21), (0,12,26), (0,16,34)</td>
</tr>
<tr>
<td>40</td>
<td>(0,1,3), (0,4,9), (0,6,20), (0,7,15), (0,10,27), (0,12,23), (0,16,34)</td>
</tr>
<tr>
<td>41</td>
<td>(0,1,10), (0,2,21), (0,3,20), (0,4,18), (0,5,16), (0,6,8), (0,19,34), (0,12,25)</td>
</tr>
<tr>
<td>44</td>
<td>(0,1,10), (0,2,21), (0,3,20), (0,4,18), (0,5,16), (0,6,13), (0,8,23), (0,12,34)</td>
</tr>
<tr>
<td>46</td>
<td>(0,1,10), (0,2,21), (0,3,20), (0,4,18), (0,5,16), (0,6,13), (0,8,23), (0,12,34)</td>
</tr>
<tr>
<td>47</td>
<td>(0,1,10), (0,2,21), (0,3,20), (0,4,18), (0,5,16), (0,6,13), (0,8,23), (0,12,34)</td>
</tr>
<tr>
<td>50</td>
<td>(0,4,9), (0,6,13), (0,8,25), (0,10,26), (0,11,18), (0,13,28), (0,2,14), (0,1,20), (0,3,23)</td>
</tr>
<tr>
<td>52</td>
<td>(0,1,3), (0,4,22), (0,5,23), (0,6,13), (0,8,25), (0,9,19), (0,12,16), (0,14,23), (0,20,11)</td>
</tr>
<tr>
<td>53</td>
<td>(0,1,3), (1,4,9), (0,6,13), (0,8,19), (0,11,23), (0,14,24), (0,16,17), (0,10,18), (0,21,10)</td>
</tr>
<tr>
<td>54</td>
<td>(0,1,3), (0,4,26), (0,6,23), (0,8,18), (0,11,27), (0,14,25), (0,5,24), (0,7,20), (0,9,21)</td>
</tr>
<tr>
<td>56</td>
<td>(0,1,3), (0,4,9), (0,6,13), (0,8,18), (0,11,26), (0,14,27), (0,16,22), (0,12,24), (0,2,23), (0,3,28)</td>
</tr>
<tr>
<td>58</td>
<td>(0,1,3), (0,4,10), (0,5,27), (0,8,17), (0,11,24), (0,14,29), (0,16,22), (0,18,21), (0,23,7), (0,12,21)</td>
</tr>
<tr>
<td>59</td>
<td>(0,1,19), (0,2,28), (0,3,7), (0,5,25), (0,8,17), (0,10,22), (0,6,21), (0,14,29), (0,24,27), (0,11,25)</td>
</tr>
<tr>
<td>60</td>
<td>(0,1,3), (0,4,9), (0,11,23), (0,7,29), (0,6,14), (0,13,28), (0,10,25), (0,24,30), (0,16,27), (0,19,21)</td>
</tr>
<tr>
<td>62</td>
<td>(0,0,1,28), (0,2,24), (0,4,19), (0,6,13), (0,8,18), (0,11,23), (0,14,30), (0,5,31), (0,9,29), (0,3,28), (0,21,24)</td>
</tr>
<tr>
<td>64</td>
<td>(0,4,9), (0,6,13), (0,8,18), (0,11,23), (0,14,29), (0,19,43), (0,17,44), (0,22,48), (0,25,28), (0,30,31), (0,2,34)</td>
</tr>
<tr>
<td>65</td>
<td>(0,1,6), (0,3,19), (0,4,26), (0,7,30), (0,8,17), (0,10,24), (0,11,24), (0,14,29), (0,18,27), (0,21,12), (0,2,25)</td>
</tr>
<tr>
<td>66</td>
<td>(0,8,18), (0,11,23), (0,16,33), (0,25,13), (0,16,26), (0,7,28), (0,9,31), (0,14,29), (0,19,24), (0,30,32), (0,1,4)</td>
</tr>
<tr>
<td>68</td>
<td>(0,4,5), (0,6,13), (0,8,18), (0,11,23), (0,14,29), (0,16,33), (0,24,49), (0,20,42), (0,27,28), (0,30,32), (0,31,34), (0,21,22)</td>
</tr>
<tr>
<td>V</td>
<td>Initial Block(s)</td>
</tr>
<tr>
<td>---</td>
<td>-----------------</td>
</tr>
<tr>
<td>70</td>
<td>(0,6,13), (0,8,18), (0,11,23), (0,14,29), (0,16,33), (0,19,39), (0,21,43), (0,26,61), (0,25,30), (0,32,34), (0,1,4), (0,24,52)</td>
</tr>
<tr>
<td>71</td>
<td>(0,1,3), (0,6,13), (0,8,18), (0,11,23), (0,14,29), (0,16,33), (0,19,39), (0,21,43), (0,27,31), (0,25,30), (0,26,35), (0,24,58)</td>
</tr>
<tr>
<td>72</td>
<td>(0,1,3), (0,6,13), (0,8,18), (0,11,23), (0,14,29), (0,16,33), (0,19,40), (0,20,45), (0,22,46), (0,28,37), (0,31,36), (0,30,34)</td>
</tr>
<tr>
<td>74</td>
<td>(0,6,13), (0,8,18), (0,11,23), (0,14,29), (0,16,33), (0,19,39), (0,21,47), (0,22,46), (0,25,26), (0,30,32), 36,73 (0,31,34)</td>
</tr>
<tr>
<td>76</td>
<td>(0,1,3), (0,6,13), (0,8,18), (0,11,23), (0,14,29), (0,16,33), (0,19,39), (0,21,46), (0,22,48), (0,27,36), (0,32,34), (0,35,38), (0,1,5), (0,24,55)</td>
</tr>
<tr>
<td>77</td>
<td>(0,1,8), (0,4,9), (0,8,18), (0,11,23), (0,14,29), (0,16,33), (0,19,39), (0,21,46), (0,24,50), (0,28,34), (0,30,32), (0,36,74), (0,37,75)</td>
</tr>
<tr>
<td>78</td>
<td>(0,1,3), (0,6,13), (0,8,18), (0,11,23), (0,14,29), (0,16,33), (0,19,39), (0,21,46), (0,22,48), (0,36,71), (0,25,52), (0,28,31), (0,32,34), (0,40,81), (0,37,75)</td>
</tr>
<tr>
<td>80</td>
<td>(0,4,9), (0,6,13), (0,8,18), (0,11,23), (0,14,29), (0,16,33), (0,19,39), (0,21,46), (0,27,55), (0,20,42), (0,24,50), (0,36,39), (0,35,37), (0,1,32), (0,1,35)</td>
</tr>
<tr>
<td>82</td>
<td>(0,4,9), (0,6,13), (0,8,18), (0,11,23), (0,14,29), (0,16,33), (0,19,39), (0,24,60), (0,36,71), (0,25,52), (0,28,31), (0,32,34), (0,40,81), (0,37,75)</td>
</tr>
<tr>
<td>83</td>
<td>(0,1,40), (0,2,38), (0,3,37), (0,4,35), (0,9,29), (0,6,32), (0,7,30), (0,13,28), (0,8,27), (0,16,33), (0,5,26), (0,10,22), (0,11,25), (0,18,42)</td>
</tr>
<tr>
<td>86</td>
<td>(0,1,43), (0,2,41), (0,3,40), (0,4,38), (0,5,36), (0,6,35), (0,7,34), (0,8,33), (0,9,30), (0,10,28), (0,11,33), (0,12,25), (0,15,32), (0,6,20), (0,4,27)</td>
</tr>
<tr>
<td>88</td>
<td>(0,1,44), (0,2,42), (0,3,41), (0,4,31), (0,5,34), (0,6,36), (0,7,35), (0,8,33), (0,9,22), (0,10,29), (0,11,28), (0,12,26), (0,15,39), (0,16,37), (0,20,52)</td>
</tr>
<tr>
<td>89</td>
<td>(0,1,44), (0,2,42), (0,3,41), (0,4,30), (0,5,23), (0,6,30), (0,7,34), (0,8,33), (0,9,22), (0,10,29), (0,11,28), (0,12,26), (0,15,39), (0,16,37), (0,20,52)</td>
</tr>
<tr>
<td>92</td>
<td>(0,1,46), (0,2,44), (0,25,65), (0,14,51), (0,5,39), (0,32,26), (0,7,36), (0,8,23), (0,9,33), (0,10,31), (0,11,30), (0,12,28), (0,4,17), (0,20,42), (0,3,38), (0,18,61)</td>
</tr>
<tr>
<td>94</td>
<td>(0,4,50), (0,2,45), (0,18,59), (0,14,52), (0,33,27), (0,7,37), (0,8,36), (0,9,34), (0,10,32), (0,11,31), (0,12,29), (0,5,26), (0,3,16), (0,1,40), (0,15,34), (0,23,47)</td>
</tr>
<tr>
<td>95</td>
<td>(0,2,46), (0,4,45), (0,18,60), (0,14,52), (0,33,27), (0,7,37), (0,8,36), (0,9,34), (0,10,32), (0,11,31), (0,12,29), (0,5,26), (0,3,16), (0,1,40), (0,15,34), (0,23,47)</td>
</tr>
<tr>
<td>98</td>
<td>(0,1,49), (0,83,30), (0,73,18), (0,14,54), (0,5,42), (0,6,41), (0,7,39), (0,8,38), (0,9,36), (0,10,34), (0,11,33), (0,12,31), (0,13,29), (0,4,21), (0,2,28), (0,3,23), (0,46,93)</td>
</tr>
</tbody>
</table>

**REFERENCES**


RECURSIVE METHOD FOR AR PARAMETER IN THE PRESENCE OF OUTLIER IN POISSON NOISE

Sadaf Manzoor
Department of Statistics, Islamia College, Peshawar, Pakistan
Email: sadaf505@yahoo.com

ABSTRACT

The study of outliers is an essential part in the field of time series as the presence of one or more outliers can seriously damage the estimation in model. Outliers in time discrete Systems, have been studied by Greta M. Ljung (1993) and a test statistic for the detection of extreme observations was developed by Manzoor S. (2010). If through the suggested test-statistic it is decided that an outlier has occurred in the autoregressive model with Poisson distribution, then the particular observation \( y_t \) can be discarded and replaced by a forecast \( \hat{y}_t \) of the true value, as a remedial measure, and the forecast is made on previous known observations. A minimum variance estimate in the sense of least square has been established in this study through Kalman Filtering recursive method.

KEYWORDS

Recursive method; outliers; time discrete systems; Forecast; least square; Kalman Filtering.

1. INTRODUCTION

It is a strong opinion that the experimentalists had faced the problem of extreme observations (influential and non-influential) when the first attempt to interpret the data, was made. If we go through history, the earliest discussion on extreme observations and outliers was made by Bernoulli (1777).

Barnett and Lewis have given the following brief definition for outliers:

“An outlier in a set of data is an observation that appears to be inconsistent with the remainder of that data set”.

In almost all sciences, the problem of outlying observation(s) creates trouble and for remedial measures one has to adopt statistical techniques. A comprehensive treatment of outliers can be seen in Barnett and Lewis (1994). They have provided a list of about 100 discordance tests for the detection of outliers regarding distributional properties.

The choice of an appropriate discordance-test depends upon:

- The distribution
- Knowledge about parameters
- Number of expected outliers, and
- Type of outliers
When the models do not fulfill the assumptions of being static, time distributed models fill this gap and researchers have to take a u-turn to study all the properties from the very first stage. A number of time series distributions play the role of a backbone in stochastic process.

Box & Jenkins model has been considered in the presence of outliers with Poisson noise, and interesting results have been found in this work.

2. SHOULD OUTLIERS BE DISCARDED?

Outliers should not be discarded blindly. Investigations are very necessary to reach the root cause. Studies have shown that sometimes these alarming extreme values led to new findings and provide useful information about the process. Outlying observations that might be appropriately dropped from one study may still be useful in another with a different objective. An outlier can be created by a shift in the location prudent to report conclusions with and without the suspected outlier in the analysis.

Removing data points on the basis of statistical analysis without an assignable cause is not sufficient to throw the data away. If outliers are discarded without definite cause, obviously they will produce under/over estimation for the parameters and wrong conclusion for the variability.

3. STOCHASTIC PROCESS

The process that describe the probability structure for dynamic models for the sequence of observations over time are known as a stochastic process can be defined as a collection of random variables over certain time points.

Different types of stochastic processes, which are mostly studied contain, purely random process, Random walk, MA process, AR process, ARIMA process, ARCH process etc.

3.1 AR Process

As the name suggests, autoregressive (AR) process is such a random process, which secures regression on itself. In this process the current value of the process is expressed as a finite, linear aggregate of previous values of the process and random shock \( a_t \).

A \( p^{th} \) order autoregressive model on equally spaced time interval satisfies the following equation:

\[
z_t = \phi_1 z_{t-1} + \phi_2 z_{t-2} + \cdots + \phi_p z_{t-p} + a_t.
\]

The autoregressive operation ‘\( p \)’ can be defined on backshift operator B such that, \( \phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p \) and the condition for stationary is: \(-1 < \phi(B) < 1\). The first order autoregressive process is also known as Markov’s process.

3.2 Existence of Outlier in AR Model

In general, the least square method works as an estimate principle for the linear model with the assumption of ‘zero’ mean and constant variance. Therefore we can find an estimate for \( \sigma_T \) with the help of the formulation
\[ \hat{\sigma}_T = (X' \Sigma^{-1} X)^{-1} X' \Sigma^{-1} y. \]

It is interesting to note that outliers can thus be estimated by the difference between the observed and interpolated value,
\[ \hat{\delta}_T = y_T - \hat{z}_T. \]

The method of missing observations and outlier detection in time series generated by Gaussian autoregressive moving average process, was discussed by Ljung (1993). Manzoor S. (2010) has developed the procedure for the noise with Poisson distribution, with following assumptions:
- The parameters of AR (1) model are known
- The time ‘T’ for expected outlier is known in advance
- The noise ‘\( \pi \)' follows Poisson distribution with parameter \( \lambda \).

The test function in the above case is:
\[ u_T = \frac{\pi_T - \phi \pi_{T+1} - (1-\phi) \lambda}{\sqrt{\lambda(1+\phi^2)}} \]

with both random variables \( \pi_T \) and \( \pi_{T+1} \) are independent, and the probability for the above function is:
\[ p(k,j) = \sum_{(q,r) \in M(k,j)} \frac{\lambda^q}{q!} \frac{\lambda^r}{r!} e^{-\lambda} \]

\[ M(k,j) = \{(q,r) \mid k-j\phi = q-r\phi\} \]

with the assumption that \( \phi \neq 0 \) is rational number since \( -1 < \phi < 1 \). And the discrete random variable \( u_T \) was described with the following values:
\[ \frac{k-\phi j - (1-\phi) \lambda}{\sqrt{\lambda(1+\phi^2)}} \]

and the probabilities \( p(k,j) \) with \( (j \& k = 0,1,K) \).

**4. FILTERING PROCESS**

One of the major problems in the presence of outlier is misleading results of parameters, and therefore wrong interpretations can be made. If through the suggested test-statistic it is decided that an outlier has occurred, then the particular observation \( y_t \) can be discarded and replaced by a forecast \( \hat{y}_t \) of the true value, as a remedial measure, and this forecast is made on previous known observations.

If such a model is available which produce \( y_t \), some Filtering Process can be used. For the states:
\[ z_t = \phi z_{t-1} + a_t \]
with no information is given except the observations, i.e,
Recursive method for AR Parameter in the presence of outlier in Poisson noise

\[ y_t = z_t + \delta x_t. \]

To reply the following 3 questions:

1. Can we estimate \( Z_t (t \text{ is fixed}) \) with observations \( Y_1, Y_2, K, Y_t \)?

2. (Extrapolation): Can we estimate \( Z_{t+\Delta}, (t+\Delta \text{ is fixed, and } \Delta > 0) \) with Observations?

3. (Interpolation): Can we estimate \( Z_t (t \text{ fixed and } < n) \) with observations \( (Y_0, Y_{t-1}, Y_t, Y_{t+1}, L, Y_n) \)?

Kalman Bucy gives us an equational solution, which is known as “Kalman Filtering”. According to the criterion, the probability for null hypothesis of no outlier \((H_0)\) given the observations can be written as:

\[
P(H_0 / y_t, Y_t^{-1}) = C e^{-\lambda} \frac{z_t^e}{e_t!} p_0
\]

\[
\Rightarrow C e^{-\lambda} \frac{z_t^e}{(y_t - z_t)!} p_0
\]

C is the normalizing constant, and to find it we need \( z_t \), which will be estimated through observations by using Kalman Filtering Procedure. Kalman Bucy Filtering can give a good approximation for \( z_t \), it depends on previous observations, and therefore iteration method has been adapted. For the alternative hypothesis, if the distribution and its properties are unknown, and no other information is given except the observations, the best solution to decide about the variable is, on the basis of observations by using filtering that:

\[
P(H_1 / y(t), Y_t^{-1}) = C \frac{1}{x_{\text{max}} - x_{\text{min}}} (1 - p_0)
\]

is Uniform distribution, where C is the normalizing constant, Bell (1983).

In order to detect that the given observation is actually an outlier, we can compare the probabilities of null and alternative hypothesis, and can take the decision about outlying observation by following rule, if:

\[
\frac{P(H_1 / y(t), Y_t^{-1})}{P(H_0 / y(t), Y_t^{-1})} \geq 1,
\]

then there exist outlier, and if:

\[
\frac{P(H_1 / y(t), Y_t^{-1})}{P(H_0 / y(t), Y_t^{-1})} < 1,
\]

there is no outlier.

4.1 Kalman Filtering Equations

The Kalman Filter is a linear, recursive estimator which produces a minimum variance estimate in the least square sense under the assumption of white, Gaussian noise.
processes. In the textbook by Chui and Chen (1991) normality is not assumed for the variables. So we can use it for Poisson noises. If we have the following model:

$\begin{equation}
\chi_{n+1/n} = \phi_{n+1} \hat{\chi}_{n/1} + K_n (I_n)
\end{equation}$

where

$I_n = z_n - H_n \hat{x}_{n/1}

K_n = \phi_n P_n H_n' (H_n P_n H_n' + R_n)^{-1}

P_{n+1} = \phi_{n+1}(P_n - P_n H_n' (H_n P_n H_n' + R_n)^{-1} H_n P_n) + G_{n+1} Q_n G_n'

$\begin{equation}
\text{and } \hat{x}_{1/0} = 0, P_n \text{ is the error covariance matrix, and } H \text{ is an indicator function. We shall use this filtering process to estimate } z_t \text{ if we have the observations } y_t. \text{ Our model is,}
\end{equation}$

$\begin{equation}
z_{t+1} = \phi z_t + a_t
\end{equation}$

and the observations

$\begin{equation}
y_t = Hz_t + e_t.
\end{equation}$

We assume that $H = 1,$

$\Rightarrow y_t = z_t + e_t$

Now $a_t = \pi_t - \lambda,$ where $\pi_t$ are independent Poisson random variables and $t = 0, 1, K,$ similarly

$e_t = \rho_t - \mu$

$\rho_t$ is independent Poisson random variate, $a$’s and $e$’s are independent.

Since $P_0 = \sigma_a^2,$ for Poisson probabilities, applying the filtering process to have an estimate, thus the first moment is:

$\begin{align}
a_t &= \pi_t - \lambda, \\
E(a_t) &= E(\pi_t - \lambda) = 0.
\end{align}$

The second moment will be:

$\begin{align}
E(a_t)^2 &= E(\pi_t - \lambda)^2 \\
\Rightarrow \sigma_a^2 &= P_0 = \lambda(1 - \lambda).
\end{align}$

Therefore for observations, since $\rho \rightarrow P(\mu),

$\begin{align}
E(e_t) &= E(\rho - \mu) = 0, \\
\text{and the second moment,}
\end{align}$

$\begin{align}
E(e_t)^2 &= \sigma_v^2 = R_0 = \mu(1 - \mu).
\end{align}$

Now we are in the position of applying Kalman Bucy recursive filtering process:
\[ \hat{z}_{t+1} = \phi \hat{z}_t + K_t I_t \]  
(5)

where \( K_t = \phi P_t (P_t + R_t)^{-1} \), \( I_t = y_t - \hat{z}_{t/t-1} \) and

\[ P_{t+1} = \phi (P_t - P_t (P_t + R_t)^{-1} P_t) \phi + \sigma^2. \]

For \( t = 0 \):

\[ z_1 = \phi \hat{z}_0 + K_0 I_0 \]
(6)

where \( \hat{z}_0 = 0 \), \( P_0 = \lambda (1 - \lambda) \) and \( R_0 = \mu (1 - \mu) \) so,

\[ K_0 = \phi P_0 (P_0 + R_0)^{-1} = \frac{\phi \lambda (1 - \lambda)}{\lambda (1 - \lambda) + \mu (1 - \mu)} \]

and \( I_0 = y_0 \), therefore by substituting:

\[ z_1 = \phi (0) + K_0 I_0 = K_0 I_0 \]

\[ \hat{z}_1 = \frac{\phi \lambda (1 - \lambda)}{\lambda (1 - \lambda) + \mu (1 - \mu)} y_0. \]

For \( t = 1 \):

\[ z_2 = \phi \hat{z}_1 + K_1 I_1 \]
(7)

\[ P_1 = \phi (P_0 - P_0 (P_0 + R_0)^{-1} P_0) \phi \]

By putting ‘\( P \)’ and ‘\( R \)’:

\[ P_1 = \phi^2 \left[ \lambda (1 - \lambda) - \frac{\lambda^2 (1 - \lambda)}{\lambda (1 - \lambda) + \mu (1 - \mu)} \right] \]

\[ \Rightarrow P_1 = \frac{\phi^2 \lambda (1 - \lambda) \mu (1 - \mu)}{\lambda (1 - \lambda) + \mu (1 - \mu)}. \]

Now, \( K_1 = \phi P_1 (P_1 + R_1)^{-1} \) will be:

\[ K_1 = \phi \left( \phi^2 \frac{\lambda (1 - \lambda) \mu (1 - \mu)}{\lambda (1 - \lambda) + \mu (1 - \mu)} \right) \left[ \frac{\phi^2 \lambda (1 - \lambda) \mu (1 - \mu)}{\lambda (1 - \lambda) + \mu (1 - \mu)} + \mu (1 - \mu) \right]^{-1} \]

By further solving:

\[ K_1 = \frac{\phi^3 \lambda (1 - \lambda) \mu (1 - \mu)}{\lambda (1 - \lambda) + \mu (1 - \mu)} \left[ \frac{\lambda (1 - \lambda) + \mu (1 - \mu)}{\mu (1 - \mu)[(1 + \phi^2) \lambda (1 - \lambda) + \mu (1 - \mu)]} \right], \]

and finally:

\[ K_1 = \frac{\phi^3 \lambda (1 - \lambda)}{(1 + \phi^2) \lambda (1 - \lambda) + \mu (1 - \mu)} \]

and \( I_1 = y_1 - \hat{z}_1 \), will become:
Thus will be:

\[ I_1 = y_1 - \frac{\phi \lambda(1-\lambda)y_0}{\lambda(1-\lambda) + \mu(1-\mu)}. \]

Thus \( z_2 = \phi I_1 \) will be:

\[
z_2 = \phi \left[ \frac{\phi \lambda(1-\lambda)y_0}{\lambda(1-\lambda) + \mu(1-\mu)} \right] + \frac{\phi^3 \lambda(1-\lambda)}{(1+\phi^2)\lambda(1-\lambda) + \mu(1-\mu)} \left[ y_1 - \frac{\phi \lambda(1-\lambda)y_0}{\lambda(1-\lambda) + \mu(1-\mu)} \right]
\]

By solving the above expression we reach to the following result, i.e,

\[
y_0 \phi^2 \lambda(1-\lambda) \left[ 1 - \frac{\phi^2 \lambda(1-\lambda)}{(1+\phi^2)\lambda(1-\lambda) + \mu(1-\mu)} \right] + \frac{y_1 \phi^3 \lambda(1-\lambda)}{(1+\phi^2)\lambda(1-\lambda) + \mu(1-\mu)}
\]

Similarly, the above iteration method can be used to find further estimated values of \( z_t \) depending upon the observation \( y_t \). Simulation study can also be made to get quick results for the above process.

**REFERENCES**


BAYESIAN ANALYSIS OF THE ADAPTIVE TYPE-II PROGRESSIVELY
HYBRID CENSORING SCHEME IN PRESENCE OF COMPETING RISKS

Fariba Hemmati\(^1\) and Esmaile Khorram\(^2\)

\(^1\) Statistical Research and Survey Department, Central Bank of the Islamic
Republic of Iran. Email: faribahematig@yahoo.com
\(^2\) Department of Statistics, Faculty of Mathematics and Computer Science,
Amirkabir University of Technology, Tehran, Iran. Email: eskhor@aut.ac.ir

ABSTRACT

There are many situations in reliability or medical studies, where items may fail due
to one of several causes. Moreover, censoring is unavoidable in any life testing or
reliability experiment. Recently Type-II progressively hybrid censoring scheme is quite
popular in reliability test, where the experiment based on a progressive censoring,
terminates at a pre-fixed time \(T\). But in this scheme the number of failures may be
observed which may have an adverse effect on the efficiency of statistical analysis. For
the propose of increasing the efficiency of the inferential producer as well as saving the
total test time, the adaptive Type-II progressively hybrid censoring scheme is introduced,
which the number of sample size \(m\) is fixed in advanced but the experiment time allowed
to run over a threshold time \(T\). If the experimental time exceeds time \(T\) but the number
of observed failures does not reach \(m\), the number of items progressively removed from
the experiment upon failure is adapted. In this paper a competing risks model based on
exponential distributions is considered under the adaptive Type-II progressively hybrid
censoring scheme. The maximum likelihood and Bayes estimators of the distribution
parameters are obtained, and two sided Bayesian probability intervals of the parameters
are also derived. In order to explain how one can apply the theoretical results obtained,
and evaluating the efficiency of estimators, a simulation study is carried out. One real
data set is used to see how different methods work in real life situations.

KEYWORDS

Competing risks; Exponential distribution; Hybrid censoring; Maximum likelihood
estimator; Type-II progressive censoring.

1. INTRODUCTION

Hybrid censoring scheme in the context of life-testing experiments was introduced at
first by Epstein (1954), which the experiment is terminated at time \(T^* = \min\{X_{m,n}, T\}\)
where \(T \in (0, \infty)\) and \(1 \leq m \leq n\) are fixed in advance, and \(X_{m,n}\) denotes the \(m\)-th failure
time when \(n\) items are placed on a life-test. This sampling scheme is considered by
several authors, for example Gupta and Kundu (1998) and Childs et al. (2003) Kundu

There are situations in which the above hybrid censoring scheme does not allow for
units to be lost or removed from the test at points other than the final termination point.
This allowance will be desirable, as in the case of accidental breakage of experimental
units, in which the loss of test units at points other than the termination point may be unavoidable. Intermediate removal may also be desirable when a compromise between reduce time of experimentation and the observation of at least some extreme lifetimes is sought. These reasons lead us directly into the area of progressive hybrid censoring.

Kundu and Joarder (2006) and Childs et al. (2008) generalized the above mentioned hybrid censoring scheme to the case when the observed sample is progressively censored. They considered Type-II progressive hybrid censoring scheme, in which \( n \) units are placed on test with censoring scheme \((R_1, R_2, \cdots, R_m)\) and stopping time \( T^* = \min\{X_{m\text{enc}}, T\} \), where \( X_{1\text{enc}} \leq X_{2\text{enc}} \leq \cdots \leq X_{m\text{enc}} \) are the order observed failure times resulting from the progressively censored experiment and \( T \) is fixed in advance. Briefly, if \( X_{m\text{enc}} < T \) the experiment terminate at time \( X_{m\text{enc}} \) and \( m \) failures occurs; Otherwise, the experiment stops at time \( T \) and only \( J \) failures occur before time \( T \), where \( X_{J+1\text{enc}} < T < X_{J+2\text{enc}} \), and \( 0 \leq J < m \). The detailed description of the Type-II progressive hybrid censoring scheme is presented in Kundu and Joarder (2006) and Childs et al. (2008) (see also Kundu et al. (2009)). Although, in order to control the total time on test, the experiment time is fixed by the experimenter, but a very smaller than \( m \) failures (or even equal to zero) may be observed which may have an adverse effect on the efficiency of the inferential producer based on Type-II progressive hybrid censoring scheme. Therefore, it is desirable to have a model that takes into account an adaptive process.

For the purpose of increasing the efficiency of statistical analysis as well as saving the total test time, Ng et al. (2009) introduced an adjustment of Type-II progressive hybrid censoring scheme, so called adaptive Type-II progressive hybrid censoring (APHC) scheme and analyzed the data under the assumptions of exponential lifetime distribution of the experimental units. Under this scheme the number of observed failures \( m \) is fixed in advanced but the experimental time is allowed to run over the (pre-fixed) threshold time \( T > 0 \). If \( X_{m\text{enc}} < T \), the experiment stops at time \( X_{m\text{enc}} \), and we will have a usual Type-II progressive censoring scheme with the pre-fixed progressive censoring scheme \((R_1, R_2, \cdots, R_m)\). If \( X_{J+1\text{enc}} < T < X_{J+2\text{enc}} \), where \( J+1 < m \), then we adapt the number of items progressively removed from the experiment upon failure by setting \( R_{J+1} = R_{J+2} = R_{m-1} = 0 \) and \( R_m = n - m - \sum_{i=0}^{J} R_i \). Thus, the effectively applied scheme is \((R_1, \cdots, R_J, 0, \cdots, 0, n - m - \sum_{i=1}^{J} R_i)\), where \( J = \max\{j: X_{J\text{enc}} < T\} \) that is, the first observed failure time exceeding the ideal total time \( T \). Therefore, as long as the failures occur before time \( T \), the initially planned progressive scheme is applied. After passing time \( T \), we do not withdraw any items at all except for the time of the \( m \)-th failure where all remaining surviving items are removed. This determination causes to terminate the experiment as soon as possible when the \((J+1)\)-th failure time is greater than \( T \), and the total test time will not be too far away from the time \( T \). If \( T = 0 \), the scheme will lead us to the case of the conventional Type-II censoring scheme, and if \( T \to \infty \), we will have a usual progressive Type-II censoring scheme. This approach illustrates how an experimenter can control the experiment. He can decide to change the value of \( T \) as a compromise between a shorter experimental time and a higher chance to observe extreme failures. Recently Lin et al. (2009) discuss inference for the Weibull distribution with the
above mentioned progressively hybrid censoring schemes. Also, Hemmati and Khorram (2011) considered analysis of the log-normal distribution under the adaptive Type-II progressively hybrid censoring scheme.

In spite of the applicability of the adaptive progressive hybrid censoring scheme, it is well known that there are situations in reliability experiments, where items may fail due to one of several causes. The theory of competing risks has been developed by researchers how have concerned with the assessment of a specific risk in the presence of other risks. The competing risks model assumes that the data consist of a failure time and an event indicator denoting the cause of failure and in the presence of censoring an incomplete observation of the failure time necessarily precludes observation of the indicator. Kundu et al. (2004) and Pareek et al. (2009) considered the analysis of competing risks data when the data are progressively Type-II censored under the latent failure times follow the exponential and Weibull distribution, respectively. Recently, Cramer and Schmiedt (2011) discussed the same latent failure time model under the assumption of independent Lomax distributed latent failure times with the same shape and different scale parameters. The analysis of Type-II progressively hybrid censored (PHC) data in present competing risks are considered by Kundu and Joarder (2006). In this paper we discuss the adaptive Type-II progressively censored data in present competing risks.

The organization of the paper is as follows. In Section 2, we introduce the model and present the notation used throughout this paper. Maximum likelihood estimation of the unknown parameters is discussed in Section 3. The Bayes estimators of the parameters and the two-sided Bayesian probability intervals of the parameters are derived in Section 4. Section 5 gives numerical results and Section 6 concludes the paper. The proofs of the Theorems which require lengthy derivations are given in the Appendix.

2. MODEL DESCRIPTION AND NOTATIONS

Consider a lifetime experiment with \( n \in N \) identical units, where its lifetimes are described by independent and identically distributed (i.i.d) random variables \( X_1, \ldots, X_n \). Without loss of generality; we assume that there are only two causes of failure. We have \( X_i = \min\{X_{i1}, X_{i2}\} \) for \( i = 1, \ldots, n \), where \( X_{ki} \), \( k = 1, 2 \), denotes the latent failure time of the \( i \) th unit under the \( k \) th cause of failure. We assume the latent failure times \( X_{i1} \) and \( X_{i2} \) are stochastically independent for \( i = 1, \ldots, n \), and the pairs \( (X_{i1}, X_{i2}) \), \( i = 1, \ldots, n \) are i.i.d. Also, we assume that the latent failure times, \( X_{ki} \), follow the exponential distribution with the cumulative distribution functions \( F_k(x) = 1 - e^{-\lambda_k x} \), and the hazard functions \( h_k = \lambda_k \).

In presence of the adaptive Type-II progressively hybrid censoring scheme under competing risks data, we have the following observation:

\[
(X_{1:n}, \delta_1, R_1), \ldots, (X_{j:n:n}, \delta_j, R_j), (X_{J+1:n:n}, \delta_{J+1}, 0), \ldots, (X_{m-1:n:n}, \delta_{m-1}, 0), (X_{m:n:n}, \delta_m, R_m),
\]

where \( J = \max\{j : X_{j:n:n} < T\} \), \( R_m = n - m - \sum_{i=1}^{m} R_i \) and \( \delta_i \in \{1, 2, *\} \). Here, \( \delta_i = k \), \( k = 1, 2 \) means the unit \( i \) has failed at time \( X_{i:n:n} \) due to cause \( k \), while \( \delta_i = * \) means the cause of unit \( i \) to fail is unknown. Let \( I_1 = \{\delta_i : \delta_i = 1\} \), \( I_2 = \{\delta_i : \delta_i = 2\} \) and
\( I_3 = \{ \delta_i : \delta_i = * \} \), then, the random variables \( m_1 = \sum_{i=1}^{m_4} I_1(\delta_i = 1) \) and \( m_2 = \sum_{i=1}^{m_4} I_2(\delta_i = 2) \) describe the number of failures due to the first and second cause of failures, respectively, and \( m_3 = \sum_{i=1}^{m_4} I_3(\delta_i = *) \) is the number of failures having failure times but corresponding causes of failure are unknown. We denote \( m_1 + m_2 = m^* \) and therefore, \( m^* + m_3 = m \). We also assume throughout that \( m^* = m_1 + m_2 \) is fixed and not random and \( m > 0 \).

In this case we observe that \( m_k \sim \text{bin}\left(m^*, \frac{\lambda_k}{\lambda_1 + \lambda_2} \right) \), \( k = 1, 2 \). Also, using the independence of the latent failure times \( X_{1i} \) and \( X_{2i} \), \( i = 1, \ldots, n \), we obtain the probabilities

\[
\pi_k = P\left( X_{ki} \leq X_{(3-k)i} \right) = \frac{\lambda_k}{\lambda_k + \lambda_{3-k}}, \quad k = 1, 2,
\]

where \( \pi_k, \ k = 1, 2 \) is the relative risk due to \( k \)th cause.

### 3. MAXIMUM LIKELIHOOD ESTIMATORS

For a given censoring scheme \( \mathcal{R} = \{R_1, R_2, \ldots, R_J, 0, \ldots, 0, n - m - \sum_{i=0}^{J} R_i \} \), where \( J = \max\{j: X_{j,m,n} < T\} \), the likelihood function of the observed data \( (x_1, \delta_1), \ldots, (x_m, \delta_m) \) is given by

\[
L(data; \lambda_1, \lambda_2 \mid J = j) =
\]

\[
c(\mathcal{R}) \prod_{i=1}^{m} \left[ f_1(x_i) \bar{F}_2(x_i) \right]^{I(\delta_i = 2)} \left[ f_2(x_i) \bar{F}_1(x_i) \right]^{I(\delta_i = 2)} \left[ f_1(x_i) \bar{F}_2(x_i) + f_2(x_i) \bar{F}_1(x_i) \right]^{I(\delta_i = *)}
\]

\[
\times \prod_{i=1}^{j} \left[ \bar{F}_1(x_i) \bar{F}_2(x_i) \right]^{R_i} \times \left[ \bar{F}_1(x_m) \bar{F}_2(x_m) \right]^{n-m-\sum_{i=1}^{J} R_i},
\]

where \( \bar{F}_k = 1 - F_k, \ k = 1, 2, \) and

\[
c(\mathcal{R}) = n \left( n - 1 - R_1 \right) \left( n - 2 - R_1 - R_2 \right) \ldots \left( n - j - R_1 - \ldots - R_j \right)
\]

\[
\times \left( n - j + 1 - R_1 - \ldots - R_j \right) \ldots \left( n - m + 1 - R_1 - \ldots - R_j \right).
\]

Using the identity \( f_k = h_k \bar{F}_k \) for \( k = 1, 2 \) we can write

\[
L(data; \lambda_1, \lambda_2 \mid J = j) =
\]

\[
c(\mathcal{R}) \prod_{i=1}^{m} \left[ h_1(x_i) \right]^{I(\delta_i = 2)} \left[ h_2(x_i) \right]^{I(\delta_i = 2)} \left[ h_1(x_i) + h_2(x_i) \right]^{I(\delta_i = *)} \left[ \bar{F}_1(x_i) \bar{F}_2(x_i) \right]^{I(\delta_i = *)}
\]

\[
\times \prod_{i=1}^{j} \left[ \bar{F}_1(x_i) \bar{F}_2(x_i) \right]^{R_i} \times \left[ \bar{F}_1(x_m) \bar{F}_2(x_m) \right]^{n-m-\sum_{i=1}^{J} R_i}
\]

\[
= c(\mathcal{R}) \lambda_1^m \lambda_2^{m_2} (\lambda_1 + \lambda_2)^{m-m^*} \exp\left\{ - (\lambda_1 + \lambda_2) \sum_{i=1}^{m} x_i + \sum_{i=1}^{j} R_i x_i + (n-m-\sum_{i=1}^{J} R_i) x_m \right\}.
\]

(2)
Taking the logarithm of $L(data; \lambda_1, \lambda_2 \mid J = j)$ and ignoring the additive constant we obtain

$$L = m_1 \log \lambda_1 + m_2 \log \lambda_2 + (m - m^*) \log(\lambda_1 + \lambda_2) - (\lambda_1 + \lambda_2)\gamma(x, j)$$  \hspace{1cm} (3)

where

$$\gamma(x, j) = \frac{m_j}{N} \sum_{i=1}^{m_j} x_{ij} + \sum_{i=1}^{j} R_i x_{ij} + \left(n - m - \sum_{i=1}^{j} R_i\right)x_{m:n}.$$  \hspace{1cm} (4)

Equating the first derivations of $L$ to zero, we get the MLEs as in the following form

$$\hat{\lambda}_k = \frac{mm_k}{m} \gamma(x, j), \hspace{1cm} k = 1, 2.$$  \hspace{1cm} (5)

Notice that $\gamma(x, j)$ corresponds to the total time on test (TTT) and the MLEs of $\lambda_1$ and $\lambda_2$ corresponds to the situation of standard Type-II progressively censoring scheme with censoring scheme $(R_1, R_2, \ldots, R_j, 0, \ldots, 0, n - m - \sum_{i=1}^{j} R_i)$.

### 4. BAYESIAN ANALYSIS

In this section we provide the Bayes estimators and the corresponding credible intervals of the unknown parameters $\lambda_1$ and $\lambda_2$. To do that, the following additional assumptions are needed:

i) The parameters $\lambda_1$ and $\lambda_2$ behaves as independent random variables.

ii) The random variable $\lambda_k$, $k = 1, 2$, has gamma prior distribution with known shape and scale parameters $\alpha_k$ and $\beta_k$, takes the following form

$$\pi_k(\lambda_k) = \frac{\beta_k^{\alpha_k}}{\Gamma(\alpha_k)} \lambda_k^{\alpha_k-1} e^{-\beta_k \lambda_k}, \hspace{1cm} \lambda_k > 0.$$  \hspace{1cm} (6)

Using binomial expansion the likelihood function (2) can be written as

$$L(data; \lambda_1, \lambda_2 \mid J = j) = c(J) \sum_{i=0}^{m-m^*} \left(\frac{m-m^*}{i}\right) \lambda_1^{m_1+i} \lambda_2^{m_2+m-m^*-i} \exp\left\{-\left(\lambda_1 + \lambda_2\right)\gamma(x, j)\right\}.$$  \hspace{1cm} (7)

The following Theorem gives the joint posterior PDF of $\lambda_1$ and $\lambda_2$.

**Theorem 4.1**

Under the assumptions (i) and (ii), the marginal posterior PDF's of $\lambda_1$ and $\lambda_2$ are give, respectively, by

$$\pi_1(\lambda_1 \mid data, J = j) = \frac{1}{A} \sum_{i=0}^{m-m^*} \left(\frac{m-m^*}{i}\right) \frac{\Gamma(\alpha_2 + m_2 + m-m^* - i)}{(\beta_2 + \gamma(x, j))^{\alpha_2+m_2+m-m^*-i} \lambda_1^{\alpha_1+m+i+l} \alpha_1+m+i-l} \times \exp\left\{-\left(\beta_1 + \gamma(x, j)\right)\lambda_1\right\}, \hspace{1cm} 0 < \lambda_1 < \infty.$$  \hspace{1cm} (8)

and
\[ \pi_2(\lambda_2 \mid \text{data, } J = j) = \frac{1}{A} \sum_{i=0}^{m^*-m} \frac{(m-m^*)}{i} \frac{\Gamma(\alpha_1 + m_1 + i) \Gamma(\alpha_1 + m_2 + m - m^* - i)}{(\beta_1 + \gamma(x, j))^{\alpha_1+m+i}(\beta_2 + \gamma(x, j))^{\alpha_2+m_2+m-m^*-i}} \times \exp\{-\beta_2 + \gamma(x, j)\lambda_2\}, \quad 0 < \lambda_2 < \infty. \]  

(9)

where

\[ A = \sum_{i=0}^{m^*-m} \frac{(m-m^*)}{i} \frac{\Gamma(\alpha_1 + m_1 + i) \Gamma(\alpha_2 + m_2 + m - m^* - i)}{(\beta_1 + \gamma(x, j))^{\alpha_1+m+i}(\beta_2 + \gamma(x, j))^{\alpha_2+m_2+m-m^*-i}}. \]

**Proof:**

See the Appendix.

Theorem 4.2

Under the assumptions (i) and (ii) the Bayes estimator for \( \lambda_1 \) and \( \lambda_2 \) are given, respectively, by

\[ \hat{\lambda}_1 = \frac{1}{A} \sum_{i=0}^{m^*-m} \frac{(m-m^*)}{i} \frac{\Gamma(\alpha_1 + m_1 + i + 1) \Gamma(\alpha_2 + m_2 + m - m^* - i)}{(\beta_1 + \gamma(x, j))^{\alpha_1+m_i+i+1}(\beta_2 + \gamma(x, j))^{\alpha_2+m_2+m-m^*-i}}, \]

(10)

and

\[ \hat{\lambda}_2 = \frac{1}{A} \sum_{i=0}^{m^*-m} \frac{(m-m^*)}{i} \frac{\Gamma(\alpha_1 + m_1 + i) \Gamma(\alpha_2 + m_2 + m - m^* - i + 1)}{(\beta_1 + \gamma(x, j))^{\alpha_1+m_i+i}(\beta_2 + \gamma(x, j))^{\alpha_2+m_2+m-m^*-i+1}}, \]

(11)

where \( A \) is defined in Theorem 4.1.

**Proof:**

See the Appendix.

Interestingly, when \( \alpha_1 = \beta_1 = \alpha_2 = \beta_2 = 0 \), the Bayes estimators coincide with the corresponding MLEs.

**Corollary 4.1:**

The minimum posterior risk associated with \( \hat{\lambda}_k \), \( k = 1, 2 \), is

\[ R_k = \mu_k^{(2)} - (\mu_k^{(1)})^2 \]

(12)

where

\[ \mu_k^{(h)} = \frac{1}{A} \sum_{i=0}^{m^*-m} \frac{(m-m^*)}{i} \frac{\Gamma(\delta_{k1}h + \alpha_1 + m_1 + i) \Gamma(\delta_{k2}h + \alpha_2 + m_2 + m - m^* - i)}{(\beta_1 + \gamma(x, j))^{\delta_{k1}h+m_i+i}(\beta_2 + \gamma(x, j))^{\delta_{k2}h+m_2+m-m^*-i}}. \]

(13)

is the marginal posterior \( h \) th moments of \( \lambda_k \), \( k = 1, 2 \), \( h = 1, 2 \), and \( \delta_{kl} = 1 \) if \( k = l \), and \( \delta_{kl} = 0 \) if \( k \neq l, l = 1, 2 \).
Proof:

See the Appendix.

Two Sides Bayesian Probability Intervals

The \(100(1-\alpha)\%\) two sided Bayesian probability intervals of \(\lambda_k, k=1,2\), say \((u_k,v_k)\), can be derived by solving the following two equations, with respect \(u_k\) and \(v_k\):

\[
\frac{\alpha}{2} = \int_0^{u_k} \pi_k(\lambda_k \mid \text{data, } J = j) d\lambda_k, \tag{14}
\]

\[
1 - \frac{\alpha}{2} = \int_0^{v_k} \pi_k(\lambda_k \mid \text{data, } J = j) d\lambda_k. \tag{15}
\]

Substituting from (8) into (14) and (15), the \(100(1-\alpha)\%\) two sided Bayesian probability interval of \(\lambda_1\) can be obtained as the solutions of the following equations with respect \(u_1\) and \(v_1\):

\[
\frac{\alpha}{2} = \frac{1}{A} \sum_{i=0}^{m-m^*} A_i \Gamma(\alpha_1 + m_1 + i, (\beta_1 + \gamma(x, j))u_1), \tag{16}
\]

\[
1 - \frac{\alpha}{2} = \frac{1}{A} \sum_{i=0}^{m-m^*} A_i \Gamma(\alpha_1 + m_1 + i, (\beta_1 + \gamma(x, j))v_1), \tag{17}
\]

where

\[
A_i = \binom{m-m^*}{i} \frac{\Gamma(\alpha_1 + m_1 + i) \Gamma(\alpha_2 + m_2 + m-m^* - i) \Gamma(\beta_2 + \gamma(x, j)) \Gamma(\alpha_2 + m_2 + m-m^* - i)}{(\beta_1 + \gamma(x, j))^{\alpha_1+m_1+i} (\beta_2 + \gamma(x, j))^{\alpha_2+m_2+m-m^* - i}}, \tag{18}
\]

and \(\Gamma(a, x)\) is the incomplete gamma function defined by

\[
\Gamma(a, x) = \frac{1}{\Gamma(a)} \int_0^x t^{a-1} e^{-t} dt.
\]

Equations (16) and (17) do not yield explicit solutions for \(u_1\) and \(v_1\) and have to be solved numerically to obtain \((u_1,v_1)\).

Similarly, from (9) into (14) and (15), we can obtain the \(100(1-\alpha)\%\) two-sided Bayesian probability interval, \((u_2,v_2)\), by solving numerically the following equations with respect to \(u_2\) and \(v_2\):

\[
\frac{\alpha}{2} = \frac{1}{A} \sum_{i=0}^{m-m^*} A_i \Gamma(\alpha_2 + m_2 + m-m^* - i, (\beta_2 + \gamma(x, j))u_2), \tag{19}
\]

\[
1 - \frac{\alpha}{2} = \frac{1}{A} \sum_{i=0}^{m-m^*} A_i \Gamma(\alpha_2 + m_2 + m-m^* - i, (\beta_2 + \gamma(x, j))v_2), \tag{20}
\]

where \(A_i\) and \(\Gamma(a, x)\) are as defined earlier.
If \( m - m^* = 0 \), the posterior density function of \( \hat{\lambda}_k, k = 1,2, \) is \( \text{Gamma}(\alpha_k + m_k, \beta_k + \gamma(x, j)) \), then the 100(1 - \( \alpha \))\% two side Bayesian probability interval of \( \lambda_k, k = 1,2 \) can be obtained using Gamma distribution, is the solutions of the following equations:

\[
\frac{\alpha}{2} = \int_0^{\nu_k} \frac{(\beta_k + \gamma(x, j))^{\alpha_k + m_k} \lambda_k^{\alpha_k + m_k - 1} e^{-(\beta_k + \gamma(x, j))d\lambda_k}}{\Gamma(\alpha_k + m_k)} \, d\lambda_k, \tag{21}
\]

\[
1 - \frac{\alpha}{2} = \int_0^{\nu_k} \frac{(\beta_k + \gamma(x, j))^{\alpha_k + m_k} \lambda_k^{\alpha_k + m_k - 1} e^{-(\beta_k + \gamma(x, j))d\lambda_k}}{\Gamma(\alpha_k + m_k)} \, d\lambda_k. \tag{22}
\]

In this case if \( \alpha_k, k = 1,2, \) is an integer, the 100(1 - \( \alpha \))\% Bayesian credible interval is

\[
\left\{ \frac{\chi^2_{(m_k + \alpha_k), 1 - \frac{\alpha}{2}}}{2(\beta_k + \gamma(x, j))}, \frac{\chi^2_{(m_k + \alpha_k), \frac{\alpha}{2}}}{2(\beta_k + \gamma(x, j))} \right\}.
\]

If no prior information is available, the non-informative priors can be used to compute the credible intervals for \( \lambda_1 \) and \( \lambda_2 \) using the above equations.

### 5. NUMERICAL RESULTS

In this Section, we report the obtained results of a simulation study, which was carried out by software R, to illustrate the theoretical results obtained in the previous sections. This simulation has done by considering different values of \( n, m \) and \( T \), and by choosing \( \lambda_1 = 0.6 \) and \( \lambda_2 = 0.8 \) in all the cases. We have used four progressive censoring schemes as

- **Scheme 1**: \( R_1 = \cdots = R_{m-1} = 0 \) and \( R_m = n - m \),
- **Scheme 2**: \( R_1 = n - m \) and \( R_2 = \cdots = R_m = 0 \),
- **Scheme 3**: \( R_1 = \cdots = R_{m-1} = 1 \) and \( R_m = n - 2m + 1 \).

For each case the maximum likelihood (ML) and Bayes estimators as described before, are computed based on 1000 simulations. It is assumed that \( m - m^* = 1 \) is fixed, and parameters \( \lambda_1 \) and \( \lambda_2 \) behaves as random variables with Gamma prior distributions with parameters (0.1,0.1) and (0.5,0.5), respectively. The average values and mean squared errors (MSEs) of the maximum likelihood and Bayes estimators of \( \lambda_1 \) and \( \lambda_2 \) for some selected choices of \( n, m \) and \( T \), have reported in Table 1. We report the average 95\% confidence intervals and the coverage probabilities of the Bayes estimates of \( \lambda_1 \) and \( \lambda_2 \) under various censoring schemes of the APHC scheme for different Ts. These results are reported in Table 2. From Table 1, it is observed that the Bayes estimators are closer to the initial values than the maximum likelihood estimators (MLEs). Also, the MSEs of
the Bayes estimators are always give the smaller MSEs in compare to those based on the MLEs. Therefore, considering the MSEs of estimators, the Bayes estimator gives better performance in estimation. As T increases, the MSEs of the ML and Bayes estimators increase for all cases. As a matter of fact, for fixed n and T, when m increases, difference between the values of estimators and initial values, also MSEs of the MLEs and Bayes estimators decrease based on the APHC scheme except for the cases of the MLEs when scheme 2 is applied. It can be due to the fact that as m increases for fixed T, some additional information is gathered. From Table 2, we observed that the coverage probabilities of the scheme 2 are always larger than those of scheme 2 and 3. In fact, what we can say is the effect of the heavy censoring at the beginning of the experiment will be affected by the number of failures and the pre-fixed threshold T.

Example:

Pareek et al. (2009) and Cramer and Schmiedt (2011) analyze progressively censored mortality data generated from a data set originally reported by Hoel (1972). It was obtained from a laboratory experiment in which male mice received a radiation does of 300 roentgens. The cause of death for each mouse was determined by autopsy. Restricting the analysis to two causes of death (reticulum cell sarcoma=cause 1; other causes=cause 2), n = 77 observations remain in the analysis. The progressively Type-II censored data was generated and first used by Kundu et al. (2004) who assume underlying exponential distributions. The censoring scheme is given by m = 25 and \( R_1 = R_2 = \cdots = R_{24} = 2, \ R_{25} = 4 \). The progressively Type-II censored sample used by Kundu et al. (2004), Pareek et al. (2009), and Cramer and Schmiedt (2011) is given by (40,2), (42,2), (62,2), (2(1,2), (206,2), (228,2), (252,2), (259,2), (318,1), (385,2), (407,2), (420,2), (462,2), (517,2), (517,2), (524,2), (525,1), (558,1), (536,1), (605,1), (612,1), (620,2), (621,1).

Here considering \( T = 550 \), so we have \( m - m^* = 0, \ m_1 = 2, \ m_2 = 17, \ j=19, \) and \( \gamma(x,j) = 29430 \). Therefore, using the exponential lifetime distribution, the maximum likelihood estimators for \( \lambda_1 \) and \( \lambda_2 \) are \( \hat{\lambda}_1 = 8.94 \times 10^{-5} \) and \( \hat{\lambda}_2 = 7.6 \times 10^{-4} \) and the Bayes estimators are \( \hat{\lambda}_1 = 1.1 \times 10^{-4} \) and \( \hat{\lambda}_2 = 7.7 \times 10^{-4} \). The maximum likelihood estimators of the relative risk rate due to cause 1 is \( \hat{\pi}_1 = 0.1053 \), and due to cause 2 is \( \hat{\pi}_2 = 0.8947 \). Also, the 95% Confidence intervals of the Bayes estimators for \( \lambda_1 \) and \( \lambda_2 \) are (-0.00071, 0.00071) and (5.0570 \times 10^{-5}, 0.00012), respectively.

6. CONCLUSION

In this paper we obtained the maximum likelihood estimators and the Bayes estimator of the parameters of the exponential distribution with the adaptive Type-II progressive hybrid censoring scheme in presence of competing risks. We have also obtained the 95% confidence intervals and coverage probabilities of the unknown parameters. We presented numerical simulation studies in order to explain how the theoretical results obtained can be applied. At last, a numerical example is presented to illustrate the methods of inference discussed in the article.
REFERENCES


APPENDIX

Proof of Theorem 4.1:

Based on the assumptions (i) and (ii) in Section 4, we obtain the joint prior PDF of $\lambda_1$ and $\lambda_2$, we obtain the joint posterior PDF of $\lambda_1$ and $\lambda_2$ as

$$
\pi(\lambda_1, \lambda_2 | data, J = j) = \frac{\pi(\lambda_1, \lambda_2) L(data; \lambda_1, \lambda_2 | J = j)}{\int_0^\infty \int_0^\infty \pi(\lambda_1, \lambda_2) L(data; \lambda_1, \lambda_2 | J = j) d\lambda_1 d\lambda_2}.
$$

(23)

On the other hand we obtain the joint PDF of $\lambda_1$ and $\lambda_2$ as

$$
\pi(\lambda_1, \lambda_2) = \frac{\beta_1 \beta_2^{\alpha_1} \alpha_2}{\Gamma(\alpha_1)\Gamma(\alpha_2)} \lambda_1^{\alpha_1-1} \lambda_2^{\alpha_2-1} e^{-\beta_1 \lambda_1} e^{-\beta_2 \lambda_2}, 0 < \lambda_1, \lambda_2 < \infty.
$$

(24)

Based on (24), the joint posterior PDF of $\lambda_1$ and $\lambda_2$ given the data and $J = j$ is

$$
\pi(\lambda_1, \lambda_2 | data, J = j) = \frac{1}{A} \sum_{i=0}^{m-m^*} \binom{m-m^*}{i} \lambda_1^{\alpha_1 + m_i + i - 1} \lambda_2^{\alpha_2 + m_2 + m - m^* - i - 1}
$$

(25)

where $0 < \lambda_1, \lambda_2 < \infty$ and

$$
A = \sum_{i=0}^{m-m^*} \binom{m-m^*}{i} \frac{\Gamma(\alpha_1 + m_i + i)\Gamma(\alpha_2 + m_2 + m - m^* - i)}{(\beta_1 + \gamma(x, j))^{\alpha_1 + m_i + i} (\beta_2 + \gamma(x, j))^{\alpha_2 + m_2 + m - m^* - i}}.
$$

According to the relation between the joint and marginal PDFs given by

$$
\pi_1(\lambda_1 | data, J = j) = \int_0^\infty \pi(\lambda_1, \lambda_2 | data, J = j) d\lambda_2,
$$

and

$$
\pi_2(\lambda_2 | data, J = j) = \int_0^\infty \pi(\lambda_1, \lambda_2 | data, J = j) d\lambda_1,
$$

we obtain the marginal posterior PDFs of $\lambda_1$ and $\lambda_2$ as

$$
\pi_1(\lambda_1 | data, J = j) =
\frac{1}{A} \sum_{i=0}^{m-m^*} \binom{m-m^*}{i} \frac{\Gamma(\alpha_2 + m_2 + m - m^* - i)}{(\beta_2 + \gamma(x, j))^{\alpha_2 + m_2 + m - m^* - i}} \lambda_1^{\alpha_1 + m_i + i - 1} e^{-(\beta_1 + \gamma(x, j))\lambda_1},
$$

(26)

$$
\pi_2(\lambda_2 | data, J = j) =
\frac{1}{A} \sum_{i=0}^{m-m^*} \binom{m-m^*}{i} \frac{\Gamma(\alpha_1 + m_i + i)}{(\beta_1 + \gamma(x, j))^{\alpha_1 + m_i + i}} \lambda_2^{\alpha_2 + m_2 + m - m^* - i - 1} e^{-(\beta_2 + \gamma(x, j))\lambda_2},
$$

(27)

where $0 < \lambda_1 < \lambda_2 < \infty$. 

Proof of Corollary 4.1:

The marginal posterior $h$ th moment of $\lambda_k$ given $J = j$, is

$$\mu_k^{(h)} = \int_0^\infty \lambda_k \pi_k(\lambda_k \mid \text{data, } J = j) d\lambda_k.$$  \hspace{1cm} (28)

For $k = 1$, on substituting (8) into (28) we can write

$$\mu_1^{(h)} = \frac{1}{A} \sum_{i=0}^{m-m^*} \frac{(m-m^*)}{i} \frac{\Gamma(\alpha_2 + m_2 + m - m^* - i)}{(\beta_2 + \gamma(x, j))} e^{-\lambda_k (\beta_1 + \gamma(x, j))} \int_0^\infty \lambda_1^{h+\alpha_1 + m_1 + i-1} \lambda_1 d\lambda_1$$

$$= \frac{1}{A} \sum_{i=0}^{m-m^*} \frac{(m-m^*)}{i} \frac{\Gamma(\alpha_2 + m_2 + m - m^* - i)\Gamma(h + \alpha_1 + m_1 + i)}{(\beta_2 + \gamma(x, j))^{h+\alpha_1 + m_1 + i}}.$$  \hspace{1cm} (29)

Similarly, for $k = 2$, replacing (28) by (9) yields

$$\mu_2^{(h)} = \frac{1}{A} \sum_{i=0}^{m-m^*} \frac{(m-m^*)}{i} \frac{\Gamma(\alpha_1 + m_1 + i)\Gamma(\delta_{k1} h + \alpha_2 + m_2 + m - m^* - i)}{(\beta_1 + \gamma(x, j))^{\alpha_1 + m_1 + i}}.$$  \hspace{1cm} (30)

Combining (29) and (30), we obtain

$$\mu_k^{(h)} = \frac{1}{A} \sum_{i=0}^{m-m^*} \frac{(m-m^*)}{i} \frac{\Gamma(\delta_{kl} h + \alpha_1 + m_1 + i)\Gamma(\delta_{k2} h + \alpha_2 + m_2 + m - m^* - i)}{(\beta_1 + \gamma(x, j))^{\delta_{kl} h + \alpha_1 + m_1 + i}}.$$  \hspace{1cm} (31)

where $\delta_{kl} = 1$ if $k = l$, otherwise $\delta_{kl} = 0$. Now,

$$R_k = \text{var}(\lambda_k \mid \text{data, } J = j) = E(\lambda_k^2 \mid \text{data, } J = j)$$

$$- E(\lambda_k \mid \text{data, } J = j)^2 = \mu_k^{(2)} - (\mu_k^{(1)})^2,$$  \hspace{1cm} (32)

which completes the proof.

Proof of Theorem 4.2:

The Bayes estimator for $\lambda_k$, $k = 1, 2$, given $J = j$, is

$$\hat{\lambda}_k = E(\lambda_k \mid \text{data, } J = j) = \mu_k^{(1)}, \ k = 1, 2.$$  \hspace{1cm} (33)

Substituting from (31) into (33) we can complete the proof of the Theorem.
Table 1:
Average values of maximum likelihood and Bayes estimates and their mean squared errors (within brackets) of the $\lambda_1$ and $\lambda_2$ under various censoring schemes of the APHC scheme for different Ts.

<table>
<thead>
<tr>
<th>T</th>
<th>(n,m)</th>
<th>Scheme</th>
<th>MLEs</th>
<th>Bayes</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>$\hat{\lambda}_1$</td>
<td>$\hat{\lambda}_2$</td>
</tr>
<tr>
<td>0.4</td>
<td>(30,10)</td>
<td>1</td>
<td>0.4242 (0.0672)</td>
<td>0.5458 (0.1125)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2</td>
<td>0.5840 (0.0910)</td>
<td>0.7610 (0.1138)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3</td>
<td>0.4404 (0.0696)</td>
<td>0.5894 (0.1050)</td>
</tr>
<tr>
<td>0.4</td>
<td>(50,10)</td>
<td>1</td>
<td>0.3433 (0.1001)</td>
<td>0.4525 (0.1639)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2</td>
<td>0.5740 (0.0856)</td>
<td>0.7901 (0.1371)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3</td>
<td>0.3551 (0.0898)</td>
<td>0.4713 (0.1488)</td>
</tr>
<tr>
<td>0.4</td>
<td>(30,15)</td>
<td>1</td>
<td>0.4648 (0.0497)</td>
<td>0.6348 (0.0694)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2</td>
<td>0.5688 (0.0530)</td>
<td>0.7433 (0.0779)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3</td>
<td>0.4919 (0.0487)</td>
<td>0.6322 (0.0713)</td>
</tr>
<tr>
<td>0.4</td>
<td>(50,15)</td>
<td>1</td>
<td>0.3681 (0.0717)</td>
<td>0.4919 (0.1155)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2</td>
<td>0.5424 (0.0531)</td>
<td>0.7204 (0.0730)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3</td>
<td>0.3551 (0.0898)</td>
<td>0.4713 (0.1488)</td>
</tr>
<tr>
<td>0.8</td>
<td>(30,10)</td>
<td>1</td>
<td>0.3375 (0.1039)</td>
<td>0.4562 (0.1577)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2</td>
<td>0.5131 (0.0749)</td>
<td>0.6654 (0.1005)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3</td>
<td>0.3603 (0.0879)</td>
<td>0.4694 (0.1467)</td>
</tr>
<tr>
<td>0.8</td>
<td>(50,10)</td>
<td>1</td>
<td>0.3380 (0.1013)</td>
<td>0.4509 (0.1667)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2</td>
<td>0.5073 (0.0687)</td>
<td>0.6671 (0.1001)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3</td>
<td>0.3439 (0.1020)</td>
<td>0.4559 (0.1744)</td>
</tr>
<tr>
<td>0.8</td>
<td>(30,15)</td>
<td>1</td>
<td>0.3633 (0.0746)</td>
<td>0.4854 (0.1223)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2</td>
<td>0.4765 (0.0544)</td>
<td>0.6400 (0.0765)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3</td>
<td>0.4021 (0.0645)</td>
<td>0.5451 (0.0974)</td>
</tr>
<tr>
<td>0.8</td>
<td>(50,15)</td>
<td>1</td>
<td>0.3343 (0.0896)</td>
<td>0.4330 (0.1580)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2</td>
<td>0.4691 (0.0540)</td>
<td>0.6381 (0.0729)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3</td>
<td>0.3349 (0.0892)</td>
<td>0.4464 (0.1508)</td>
</tr>
</tbody>
</table>
Table 2: The average 95% confidence intervals and the coverage probabilities of the Bayes estimates of $\lambda_1$ and $\lambda_2$ under various censoring schemes of the APHC scheme for different $Ts$.

<table>
<thead>
<tr>
<th>$T$</th>
<th>(n,m)</th>
<th>Scheme</th>
<th>Confidence interval</th>
<th>Coverage probability</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>$\hat{\lambda}_1$</td>
<td>$\hat{\lambda}_2$</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.4</td>
<td>(30,10)</td>
<td>1</td>
<td>(0.1287, 0.9129)</td>
<td>(0.2086, 1.0991)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2</td>
<td>(0.1768, 1.2532)</td>
<td>(0.2861, 1.4971)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3</td>
<td>(0.1320, 0.9526)</td>
<td>(0.2261, 1.1756)</td>
</tr>
<tr>
<td></td>
<td>(50,10)</td>
<td>1</td>
<td>(0.1041, 0.7415)</td>
<td>(0.1754, 0.9143)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2</td>
<td>(0.1701, 1.2444)</td>
<td>(0.2990, 1.5393)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3</td>
<td>(0.1068, 0.7688)</td>
<td>(0.1824, 0.9508)</td>
</tr>
<tr>
<td></td>
<td>(30,15)</td>
<td>1</td>
<td>(0.1795, 0.8963)</td>
<td>(0.2965, 1.1316)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2</td>
<td>(0.2227, 1.0870)</td>
<td>(0.3421, 1.3215)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3</td>
<td>(0.1932, 0.9392)</td>
<td>(0.2922, 1.1343)</td>
</tr>
<tr>
<td></td>
<td>(50,15)</td>
<td>1</td>
<td>(0.1430, 0.7088)</td>
<td>(0.2308, 0.8869)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2</td>
<td>(0.2113, 1.0396)</td>
<td>(0.3332, 1.2801)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3</td>
<td>(0.1068, 0.7688)</td>
<td>(0.1824, 0.9508)</td>
</tr>
<tr>
<td>0.8</td>
<td>(30,10)</td>
<td>1</td>
<td>(0.1012, 0.7324)</td>
<td>(0.1778, 0.9187)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2</td>
<td>(0.1556, 1.1021)</td>
<td>(0.2521, 1.3219)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3</td>
<td>(0.1089, 0.7778)</td>
<td>(0.1811, 0.9497)</td>
</tr>
<tr>
<td></td>
<td>(50,10)</td>
<td>1</td>
<td>(0.1017, 0.7320)</td>
<td>(0.1751, 0.9096)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2</td>
<td>(0.1523, 1.0932)</td>
<td>(0.2529, 1.3229)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3</td>
<td>(0.1042, 0.7435)</td>
<td>(0.1768, 0.9187)</td>
</tr>
<tr>
<td></td>
<td>(30,15)</td>
<td>1</td>
<td>(0.1415, 0.6991)</td>
<td>(0.2279, 0.8752)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2</td>
<td>(0.1850, 0.9161)</td>
<td>(0.2980, 1.1418)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3</td>
<td>(0.1557, 0.7751)</td>
<td>(0.2556, 0.9776)</td>
</tr>
<tr>
<td></td>
<td>(50,15)</td>
<td>1</td>
<td>(0.1313, 0.6404)</td>
<td>(0.2027, 0.7854)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2</td>
<td>(0.1817, 0.9035)</td>
<td>(0.2979, 1.1371)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3</td>
<td>(0.1306, 0.6443)</td>
<td>(0.2101, 0.8063)</td>
</tr>
</tbody>
</table>
ROOT TRANSFORMATION METHOD FOR MONITORING CORRELATED VARIABLE AND ATTRIBUTE QUALITY CHARACTERISTICS

Mohammad H. Doroudyan and Amirhossein Amiri
Department of Industrial Engineering, Shahed University, Tehran, Iran
Email: doroudyan@shahed.ac.ir, amiri@shahed.ac.ir

ABSTRACT

Nowadays, quality of many products or processes is represented by two or more correlated quality characteristics. Hence, many control charts have been developed to monitor these quality characteristics in types of multivariate and multi-attribute quality characteristics, separately. However, sometimes quality of a product or a process can be characterized by combination of correlated variable and attribute quality characteristics. To the best of our knowledge there is no research in this area. In this paper a method is proposed to monitor these quality characteristics based on root transformation technique. In the proposed method we first transform the distribution of correlated variable and attribute quality characteristics to approximate multivariate normal distribution based on the reduction of skewness. Then, multivariate control charts such as $T^2$ are used to monitor transformed data. The performance of the proposed method is evaluated under different step shifts by using simulation studies in terms of average run length criterion. The results show that the proposed method performs satisfactory.

KEYWORDS

Statistical process control (SPC); correlated variable and attribute quality characteristics; root transformation; skewness reduction; phase II; average run length (ARL).

1. INTRODUCTION

Control charts are the most important tools of statistical process control (SPC) used to distinguish the assignable and common causes of variation and leads to improvement in the process. The first control chart is introduced by Shewhart (1931). After that, various control charts are developed by many authors for monitoring different quality characteristics under different situations. For example, refer to exponentially weighted moving average (EWMA) control chart by Roberts (1959) and cumulative sum (CUSUM) chart by Page (1954) in monitoring a variable quality characteristic. Also, see Woodall (1997) for a review on monitoring attribute quality characteristics.

Hotelling (1947) showed that monitoring two or more correlated quality characteristics separately leads to increasing in probability of Type I error in control charts. Type I error occurs when the process is in-control state and control chart gives an alarm. Therefore, researchers have developed control charts for monitoring multivariate and multi-attribute quality characteristics, separately. For example, see some of important approaches in monitoring multivariate quality characteristics in Bersimis et al. (2007). Topalidou and Psarakis (2009) have also reviewed multi-attribute control charts.
In some cases, the quality of a product or a process is represented by correlated variable and attribute quality characteristics. For instance, in plastic manufacturing companies, number of nonconformance and weight of the product are correlated. Since monitoring these quality characteristics individually leads to misleading results, these quality characteristics should be monitored simultaneously. To the best of our knowledge, there is no method in monitoring these types of quality characteristics. Only in one research confidence bound of defect rate is obtained by Kang and Brenneman (2010) for attribute and variable data, but they considered independent assumption for these quality characteristics. In this paper, we propose a method to monitor correlated variable and attribute quality characteristics. Based on the proposed method, we first transform original data by root transformation method which is proposed by Niaki and Abbasi (2007) in such a way that the marginal distribution of transformed data is approximate normal. We then apply multivariate $T^2$ control chart to monitor the transformed data. The performance of the proposed method is evaluated under different step shifts in mean vector of quality characteristics by using simulation studies in terms of average run length (ARL) criterion.

In the next section the problem is defined and formulated. We explain the proposed method in section 3. The performance of the root transformation method is evaluated in section 4. And finally, section 5 includes comments and some suggestions for future research.

2. PROBLEM DEFINITION

In this paper, we consider a condition in which the quality of a product or a process is represented by the combination of correlated variable and attribute quality characteristics in the $X = (x_1, x_2)^T$ vector where $x_1$ is variable and follows normal distribution and $x_2$ is attribute quality characteristic and occurs on Poisson distribution. Also, we are in phase II of control chart, therefore, the mean vector ($\mu_X$) and covariance matrix ($\Sigma_X$) of joint distribution of $X$ are known based on historical data. Since shift in the mean of Poisson distribution leads to changes the variance of it, we assume that the correlation between these quality characteristics is stable. The purpose of this paper is monitoring the mean vector ($\mu_X$) of quality characteristics.

3. PROPOSED APPROACH

As explained in the introduction section, due to correlation between correlated variable and attribute quality characteristics, monitoring each quality characteristic separately leads to increasing probability of Type I error and misleading results. In this paper, we propose root transformation method to transform quality characteristics in such a way that the joint distribution of them follows bivariate normal distribution, approximately. Then, a $T^2$ multivariate control chart is used to monitor the transformed data. Note that, several transformation methods have been proposed for this purpose. For example, refer to arcsin transformation by Anscombe (1948), root square transformation by Ryan (1989), Q-transformation by Quesenberry (1997), and recently, root transformation by Niaki and Abbasi (2007). It is shown that the root transformation method performs better than the other methods under monitoring multi-attribute quality characteristics.
characteristics (Niaki and Abbasi, 2007). Hence, we used this transformation method in our paper.

Based on the root transformation method, for \( i^{th} \) quality characteristic, \( r_i \) is determined in such a way that the marginal distribution of \( r^{th} \) root transformation \( (y_i=x_i') \) has zero skewness. Moreover, the mean vector \( (\mu_{Y}) \) and covariance matrix \( (\Sigma_{Y}) \) of transformed data \( (Y=(y_1, y_2)^T) \) change and are estimated based on the transformed historical data using moment method.

In order to find appropriate transformation, a bisection method is employed based on historical data to find the power \( (r) \) of the root transformation for each quality characteristic.

The bisection method is based on the opposite signs of function in two side of initial interval. The sign of function is evaluated at the central point of an interval and the central point is replaced instead of limit which has the same sign. The bisection method will be continued until the value of function becomes less than predefined value.

For example, to find a root of \( f(r) = 0 \) in the interval of \((a_0, b_0)\) where \( f(a_0)f(b_0) < 0 \), we define desired stopping value \( (\varepsilon) \) and then apply the following algorithm:

\[
k = 0, \\
\text{while } |f(r_{k+1})| > \varepsilon \\
\quad r_{k+1} = (a_k + b_k)/2 \\
\quad \text{if } (f(r_{k+1})f(a_k) < 0), \text{ then} \\
\quad \quad a_{k+1} = a_k \text{ and } b_{k+1} = r_{k+1} \\
\quad \text{else} \\
\quad \quad b_{k+1} = b_k \text{ and } a_{k+1} = r_{k+1} \\
\text{end if} \\
\quad k = k+1 \\
\text{end while}
\]

\[r^* = r_k\]

In the root transformation method, we consider \( f(r) \) as the value of skewness on the \( r^{th} \) root transformation of quality characteristics \( (x_i') \). The purpose is finding \( r \) such that \( f(r) \) becomes near to zero. Therefore, we want to find a root for \( f(r) \approx 0 \) in the initial interval \((0, 1)\).

Since the variable quality characteristic has normal distribution (based on assumption) we always set \( r_1=1 \) for it. After finding a proper root \( (r_2) \) for transforming attribute data, mean vector and covariance matrix of the transformed data is estimated. In each sampling first transformed quality characteristics are calculated then we can use \( T^2 \) control chart to monitor the transformed data. The statistic of control chart is \( T^2 = n (Y - \mu_{Y})^T \Sigma_{Y}^{-1} (Y - \mu_{Y}) \) where \( n \) is the number of samples in each subgroup. The upper control limit (UCL) is
where $\alpha$ is the desired probability of Type I error in control chart. In the control chart when the control statistic becomes greater than upper control limit, it is considered as a signal. For more information about $T^2$ control chart, see Mason and Young (2002). Summarized steps are as follows:

- Find a proper root transformation by using bisection method for each quality characteristic based on historical data (Note that for variable quality characteristic the power of root transformation is always set equal to 1).
- Estimate the mean vector and covariance matrix of transformed data by using moment method.
- Calculate the transformed vector for each sample and monitor transformed data in $T^2$ control chart.

In the next section, the performance of root transformation method for correlated variable and attribute quality characteristics is evaluated by using simulation studies.

### 4. SIMULATION STUDIES

In this section, performance of the proposed method is evaluated using a numerical example. In this example, quality of a product or a process is characterized by a combination of correlated variable and attribute quality characteristics, where the variable quality characteristic follows a normal distribution with mean 3 and variance 4 and the attribute quality characteristic follows Poisson distribution with parameter 4, and the correlation between them is equal to 0.35.

In the first subsection, the performance of root transformation is investigated using normality test by Jarque and Bera (1987). In the second subsection, the performance of control chart in monitoring transformed quality characteristics is evaluated in terms of out-of-control average run length (ARL$_1$) when the mean vector changes in different scenarios.

#### 4.1 Performance of Root Transformation Method

The historical five thousand random vectors are generated using Gaussian copula (Cherubini et al. 2004). Then mean vector and covariance matrix of quality characteristics is obtained using moment method as follows:

$$\hat{\mu}_x = (3.01 \ 4.01)^T \quad \text{and} \quad \hat{\text{Cov}}(X) = \begin{pmatrix} 4.03 & 1.4 \\ 1.4 & 4.04 \end{pmatrix}.$$  

Skewness of quality characteristics are 0.03 and 0.49 for variable and attribute quality characteristics, respectively. The normality of variable quality characteristic is accepted and the normality of attribute quality characteristic is rejected based on the Jarque and Bera normality test. In the first step, we set $r_1 = 1$ for variable quality characteristic ($y_1 = x_1^1$) and find the proper root ($r_2$) for the attribute quality characteristic by using the bisection method ($y_2 = x_2^{r_2}$). The values of 1 and 0.72 are obtained for $r_1$ and $r_2$, respectively. The mean vector and covariance matrix of transformed data are estimated by using moment method as follows:

$$\hat{\mu}_y = (3.01 \ 2.63)^T \quad \text{and} \quad \hat{\text{Cov}}(Y) = \begin{pmatrix} 4.03 & 0.7 \\ 0.7 & 1.01 \end{pmatrix}.$$
The skewness values of transformed data are reduced to 0.03 and zero. The Jarque and Bera normality test approves the normality of transformed variable and attribute data with the P-values of 0.70 and 0.42, respectively.

The results show that root transformation method transforms the correlated variable and attribute quality characteristics to the approximate multivariate normal distribution.

4.2 Monitoring Transformed Data

In this subsection, the performance of the control chart in monitoring transformed correlated variable and attribute quality characteristics is evaluated using simulation studies in terms of out-of-control average run length (ARL) criterion. In order to monitor the transformed data, in each sampling of this simulation, one vector of correlated variable and attribute quality characteristics is simulated. Then, the sample vector is transformed based on proper root transformation obtained in previous subsection. And finally, the transformed data is monitored by $T^2$ control chart. Note that the mean vector and covariance matrix of transformed data which is used in control statistic are obtained in the previous subsection.

We first set the upper control limit (UCL) of the $T^2$ control chart equal to $\chi^2_{2,0.005} = 10.59$ to obtain the $ARL_0 = 200$. Then, this UCL is tested by simulation studies and $ARL_0 \approx 219$ is obtained after 10000 replications. In order to evaluate the performance of control chart in detecting out-of-control state, we consider different step shift scenarios in the mean vector of quality characteristics. Then, out-of-control state is simulated to obtain out-of-control run length, these simulations are replicated 10000 times for each scenario and the results of $ARL_1$'s are summarized in Table 1. Note that the number of samples in each subgroup is considered equal to 1 ($n=1$). Meanwhile, in each scenario the $(\theta_1, \theta_2)$ vector shows the magnitude of shift in the mean vector of quality characteristics in the unit of its standard deviation ($\sigma$). For example $(2\sigma, 2\sigma)$ shows that the mean vector of quality characteristics is changed to $(\mu_1+2\sigma, \mu_2+2\sigma)$.

The results of $ARL_1$ for different shifts considered in the paper are summarized in Table 1 as follows:

| ARL values for step shifts in the mean vector of quality characteristics |
|-----------------------------|----------------|----------------|----------------|----------------|
| $(\sigma, 0)$               | $(0, \sigma)$  | $(2\sigma, 0)$ | $(0, 2\sigma)$ | $(3\sigma, 0)$ |
| 37.98                       | 30.59           | 5.72            | 6.43           | 1.87           |
| $(\sigma, \sigma)$         | $(2\sigma, 2\sigma)$ | $(3\sigma, 3\sigma)$ | $(-\sigma, 0)$ | $(0, -\sigma)$ | $(-\sigma, -\sigma)$ |
| 23.64                       | 4.18            | 1.63            | 37.38          | 75.58          |

The results of Table 1 show that the changes in the mean vector of the original data can be detected soon in all shift scenarios considered in this paper which shows effectiveness of the proposed method in monitoring correlated variable and attribute quality characteristics.
5. CONCLUSIONS AND FUTURE RESEARCHES

In this paper, we considered a skewness reduction approach to monitor correlated variable and attribute quality characteristics. For this purpose, we proposed root transformation method to transform the correlated variable and attribute quality characteristics in such a way that the transformed data follows approximate multivariate normal distribution. We then used multivariate $T^2$ control chart for monitoring the transformed quality characteristics. The performance of the proposed method was evaluated based on average run length (ARL) criterion by using simulation studies. The results showed the effectiveness of the proposed method. As a future research, other transformation techniques can be addressed by researchers for monitoring correlated variable and attribute quality characteristics. Moreover, other multivariate control charts such as MEWMA can be used to monitor these transformed data.

REFERENCES

USING TIME DEPENDENT COEFFICIENT RATES MODEL FOR ASSESSING THE IMPACT OF BREASTFEEDING ON WEIGHT LOSS IN LESS THAN TWO YEARS OLD CHILDREN

Alireza Abadi¹, Gelareh Rahimi², Yadollah Mehrabi³ and Nahid Kholdi⁴

¹ Department of Community Medicine, Faculty of Medicine, Shahid Beheshti University of Medical Sciences, Iran. Email: alirezaabadi@gmail.com
² Fertility-Infertility Health Research Center (IRHRC), Taleghani Hospital, Shahid Beheshti University of Medical Sciences, Iran
³ Department of Epidemiology, Faculty of Public Health, Shahid Beheshti University of Medical Sciences, Iran
⁴ Department of Community Medicine, Faculty of Medicine, Shahed University, Iran

SUMMARY

Although breast milk is universally recommended as the preferred source of infant nutrition, there is little agreement regarding the optimal duration of breastfeeding. In this study we used time dependent coefficient rates model in order to investigate the association between stopping breastfeeding and weight loss in less than 2 years old children where the association varies with age at which breastfeeding stopped. Regression B-splines were used for estimating time varying coefficients in the proportional rates model. Furthermore by using a Wald type test the variability of the effect of stopping breastfeeding on weight loss were checked. We also evaluated goodness of fit of the model by applying a test based on martingale residuals. The study suggests that, weaning before 3 months increases the rate of weight loss by 10 times. This relative rate decreases to 4.7 for babies who were breastfed less than 4 months and decreases again to 3.4 for babies breastfed less than 5 months. In summary we observed a decreasing trend of relative rate of weight loss for infants who were not breastfed for determined number of months.

KEYWORDS

Regression splines; recurrent event data; time varying coefficient; proportional rates model; weight loss

INTRODUCTION

Weight loss in babies is a major health problem especially in developing countries. Infectious diseases and poor nutrition are known to play important roles in infant weight loss. Previous studies have shown that the role of malnutrition is not limited to only severe cases of malnutrition (1-3). It is clear that the greater degree of malnutrition, the worse the outcomes, but under-nutrition even in a mild form, can have detrimental effects during childhood. About one in every three people in the world is suffering from malnutrition (4). Half of the child deaths in developing countries are a result of malnutrition and weight loss (5).
Malnutrition is not uncommon especially in rural and southern Iran. According to UNICEF, in 2004, 5% of under five years old children were suffering from underweight. The last national survey in Iran in November 1998 showed that 12.8% of children under 5 years old in rural areas are moderately or severely stunted (<-2 SD weight-for-age) and about 4.8% of them are moderately or severely wasted (<-2 SD weight-for-height). Statistics are substantially different between rural and urban areas. 22 percent of rural children aged under fives are stunted compared to 11·0% of urban children (6-7).

In order to prevent the problem of underweight, first step is identifying risk factors associated with underweight and stunting in children.

Various studies have found that there are several risk factors leading to incidence of malnutrition and weight loss (8-15). To date it is know that mother’s literacy status [], family income, water connection, birth order of the child, birth weight, sex, illnesses and lack of breastfeeding are of potential risk factors for child malnutrition.

Breastfeeding has always been the gold standard for infant feeding. The benefits of breastfeeding to infants, mothers and society have been well-documented. On average, breastfed babies have fewer infections in their early life than non-breastfed babies. In particular, they have less diarrhea and vomiting, chest infections, and ear infections compared to babies who are not breast fed (16).

Observational studies have shown that malnutrition rates begin to increase substantially at the age that children are no longer exclusively breastfed (17).

Due to the fact that Breastfeeding and its effects continue to attract the attention of investigators, the authors performed a study in order to investigate the impact of stopping breastfeeding at different child age on weight loss in less than 2 years old children. In order to assess month-specific effect of not breastfeeding on children weight loss we need to allow its regression coefficient to vary with the age at which breastfeeding stopped.

Hastie and Tibshirani (1993) proposed general varying-coefficient models. In the cox model framework, it is assumed that regression coefficient is a function of the follow up time (18). In 2003 Bin et al. proposed varying coefficient Cox model which could be used to explore the functional form of the relationships between covariates and time-to-event, and also examine whether and how a particular effect is changing over time (19). They estimated the nonparametric function \( \beta(t) \) using regression spline method by approximating \( \beta(t) \) using the natural cubic B-spline basis. However this model is defined for univariate time to event data. Amorim et al. proposed a generalization of this method that is useful in estimating time varying coefficients in recurrent event data settings (20). They considered proportional rates model as a basis model and then incorporated a b-spline function into it. Rate models could be utilized to formulate the event recurrence rate as a function of observed covariates.

In this study we used time-dependent coefficient rates model with the aim of assessing stopping breastfeeding on weight loss occurrence rate in less than 2 years old children.
SUBJECT AND METHODS

Data were collected from records of babies who were visited for monthly check ups at Tehran east health centers. The analysis included 1508 new born infants who were born without birth defects. Information on each infant’s vital status and feeding mode was collected at baby doctor visits which took place every month until the infant was 11 months of age and then every two months till the baby reached 23 months. At each visit, mothers were asked whether they had offered their child to eat or drink breast milk, complementary food, vitamins and supplements during the past month (or 2 months). They were also asked if their child had experienced teething, weight loss, an episode of diarrhea, infectious diseases or any kind of illnesses. Weight loss was defined as a decrease of 50gr or more from previous month weight.

Maternal socio-demographic characteristics including mother’s age and educational level were also considered in the analysis.

We first fitted the standard proportional rates model to this data with the purpose of estimating the effect of non breastfeeding on the weight loss recurrence rate as covariate that its effect is assumed to be constant:

$$d\mu_i(t) = \exp\{\beta^T Z_i(t)\}d\mu_0(t)$$

where $d\mu_0(t)$ is the baseline rate function, $Z(t)$ is the vector of covariate values at time t, and $\beta$ is the vector of regression parameters. In our practice the time t corresponds to the age of child in months.

We then applied time-dependent coefficient rates model in order to assess the impact of stopping breastfeeding on recurrence rate of weight loss with respect to the child’s age at which breastfeeding stopped:

$$d\mu_i(t) = \exp\{\beta^T Z_i(t) + \theta(t)W_i(t)\}d\mu_0(t)$$

where $Z(t)$ is the vector of covariates that their effects are assumed to be constant over time, and w(t) is a covariate or an exposure variable that its effect is supposed to be changeful at different time points. $\theta(t)$ is the time varying regression parameter which could be estimated by standard cubic B-spline basis functions, such that:

$$\theta(t) = \gamma_0 + \sum_{k=1}^{m+3} \gamma_k \bar{B}_k(t)$$

In this equation $\bar{B}(t)$ is the cubic B-spline basis function and m is the number of interior knots. We used Akaike's Information Criterion for determining the optimal number of knots.

Socio-demographic characteristics of mother and all of the information about infant’s vital status and feeding mode except breastfeeding status were considered as covariates that their effects are presumed to be fixed at different child ages, and breastfeeding status was primary variable of interest which was thought to have time-dependent effect.

In order to formally test the hypothesis that the effect of interest changes over time a Wald type statistic were used. For the purpose of comparing the applied models we used Akaike's Information Criterion once more and then picked the model with minimum AIC as the best model.

Goodness of fit of the selected model was examined by applying a test based on martingale residuals.
RESULTS

Data of 1508 children who were born without birth defects was included in the analysis; among those children 814 (54%) were boys and 694 (46%) girls. Mean birth weight was 3219 gr (SD 448 gr). 517 of (34%) babies had experienced weight loss at least once during the first two years of life.

The effect of breastfeeding was first determined by applying proportional rates model after entering other significant prognostic factors (Table 1).

The results shows that the occurrence rate of weight loss for those babies who didn’t receive breast milk is 89% higher than breastfed babies (P<0.001; $\hat{\beta} =0.64$). In other words relative rate of weight loss in babies who breastfed less than t months versus babies who breastfed at least t month is 1.89, for all values of t (1 $\leq t \leq 23$).

On the next step time-varying coefficient rates model was fitted to estimate the month-specific effect of breastfeeding on infant’s weight loss adjusting for significant prognostic factors listed in table 1. Table 2 provides the estimated coefficients of stopping breastfeeding related to different weaning ages, considering 3 interior knots. Since the coefficient difference of other variables was shown to be negligible compared to values listed in table 1, they are not presented here once more. The results revealed that the effect of stopping breastfeeding is dependent on breastfeeding duration. Figure 1 contains the log of the relative rate of weight loss as a function of the age at which breastfeeding stopped, considering 3 interior knots. It is obvious that there is a decreasing trend of relative rate of weight loss during time. That is to say the more the duration of breastfeeding the less the occurrence rate of weight loss.

For better understanding we also plotted relative rate of weight loss against the age at stopping breastfeeding in figure 2. As we can see, lack of breastfeeding for the first four months of child’s life strongly increases the risk of weight loss. For instance the rate of weight loss occurrence for a baby who breastfed less than 3 months is approximately 6 times more than a baby who breastfed at least 3 months.

Results from Wald-type statistic suggest that the impact of breastfeeding on the occurrence of weight loss is significantly different with respect to the age at which child was no longer breastfed (P<0.001). Calculating Akaike's Information Criterion for both fitted models showed that time-dependent coefficient rates model has smaller AIC (8066) than proportional rates model (8086), therefore based on min-AIC criterion, the model with time varying coefficient for breastfeeding status is more appropriate.

Selected model’s goodness of fit was checked by a test based on martingale residuals, and the results confirmed that the model is well fitted (P=0.166).

DISCUSSION

Several investigators have reported a positive and significant association between non breastfeeding and growth faltering in children. According to Hugh et al. (2004) weight for age malnutrition rates begin to increase substantially at the age that children are no longer exclusively breastfed (17). Islam et al. (1994) have also found that lack of breastfeeding would increase the risk of malnutrition in children by 4 times (21). Another
study in 1988 revealed that diarrhea have no significant impact on the growth of exclusively breastfed infants (22). Chandra in 1981 and Diaz et al. in 1995 both showed the adequacy of breastfeeding to support infant growth (23-24).

Considering the fact that weaning age varies from one child to another, the impact of stopping breastfeeding on baby’s weight loss is supposed to be proportional to the duration of breastfeeding. Aniango et al. (1999) compared growth rate in three groups of babies where the three groups differed with respect to duration of breastfeeding (25). They reported a dose-response association between breastfeeding duration and weight gain.

Kramer et al. (2003) has also examined the effects on infant growth and health of 3 compared to 6 months of exclusive breastfeeding and showed that longer durations of breastfeeding results in lower risk of adverse health effect in the first year of life (26).

However the performed studies were rather limited, since the comparison was made among 2 or 3 groups of babies with different duration of breastfeeding. In this practice we used time-dependent coefficient rates model for estimating the month-specific effect of stopping breastfeeding on infant’s weight loss. Results of our study revealed that the impact of breastfeeding on the occurrence rate of weight loss is different with respect to the weaning age. Particularly curtailment of breastfeeding during first 3 months of child’s life critically increases the risk of weight loss. This negative effect tends to weaken as the baby grows older and completely disappears after the baby passes the one year of age.

Our study suggested that using marginal standard rate model in situations where the effect of interest tends to be unstable over time would be misleading. Fitting this model to our data gives the relative rate of stopping breastfeeding to be 1.89 regardless of length of time baby was breastfed.

Table 1:
Results of applying proportional rates model for determining the effect of non-breastfeeding on infant’s weight loss after adjusting for other prognostic factors

<table>
<thead>
<tr>
<th>Prognostic Factor</th>
<th>Relative Rate</th>
<th>95% Confidence Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>Literacy of Mother†</td>
<td>0.69</td>
<td>0.52, 0.9</td>
</tr>
<tr>
<td>Occupational Status of Mother‡</td>
<td>1.66</td>
<td>1.24, 2.22</td>
</tr>
<tr>
<td>Diarrhea</td>
<td>18.59</td>
<td>14.29, 24.17</td>
</tr>
<tr>
<td>Cold and Respiratory Infections</td>
<td>16.99</td>
<td>13.29, 21.71</td>
</tr>
<tr>
<td>Urinary Infections</td>
<td>15.81</td>
<td>8.05, 31.03</td>
</tr>
<tr>
<td>Stopping Breastfeeding</td>
<td>1.89</td>
<td>1.47, 2.43</td>
</tr>
<tr>
<td>Complementary food</td>
<td>7.93</td>
<td>6.24,10.09</td>
</tr>
</tbody>
</table>

†  Children who’s mothers had academic education in comparison to the children who’s mothers did not have academic education
‡  Children who’s mothers were occupied in comparison to children who’s mothers were housewife
Using time dependent coefficient rates model for assessing the...

Table 2:
Month specific estimation of stopping breastfeeding on infant’s weight loss

<table>
<thead>
<tr>
<th>Stopping breastfeeding before month</th>
<th>Relative rate of weight loss</th>
<th>95% confidence interval</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>11.76</td>
<td>0.70,196</td>
<td>0.086</td>
</tr>
<tr>
<td>2</td>
<td>9.69</td>
<td>3.06,30</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>3</td>
<td>6.89</td>
<td>2.51,18.92</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>4</td>
<td>4.72</td>
<td>2.13,10.49</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>5</td>
<td>3.48</td>
<td>1.98,6.12</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>6</td>
<td>2.99</td>
<td>1.82,4.92</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>7</td>
<td>2.88</td>
<td>1.89,4.39</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>8</td>
<td>2.89</td>
<td>2.00,4.17</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>9</td>
<td>2.83</td>
<td>1.93,4.13</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>10</td>
<td>2.55</td>
<td>1.73,3.76</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>11</td>
<td>2.15</td>
<td>1.49,3.12</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>13</td>
<td>1.41</td>
<td>0.89,2.25</td>
<td>0.14</td>
</tr>
<tr>
<td>15</td>
<td>1.01</td>
<td>0.56,1.84</td>
<td>0.96</td>
</tr>
<tr>
<td>17</td>
<td>0.94</td>
<td>0.57,1.56</td>
<td>0.81</td>
</tr>
<tr>
<td>19</td>
<td>1.06</td>
<td>0.65,1.73</td>
<td>0.81</td>
</tr>
<tr>
<td>21</td>
<td>1.25</td>
<td>0.77,2.03</td>
<td>0.36</td>
</tr>
<tr>
<td>23</td>
<td>1.35</td>
<td>0.70,2.63</td>
<td>0.37</td>
</tr>
</tbody>
</table>

Fig. 1: Estimated log(RR) for the weight loss occurrence as a function of age at which breastfeeding stopped (solid curves) and corresponding pointwise 95% CI (vertical bars) at selected ages. Estimates from proportional rates model (dotted line) and its 95% CI (dashed lines) are also presented.
Fig. 2: Estimated RR for the weight loss occurrence as a function of age at which breastfeeding stopped (solid curve). Estimates from proportional rates model are shown with dotted line.

REFERENCES

A DATA ENVELOPMENT ANALYSIS METHOD FOR OPTIMIZING MULTI-RESPONSE PROBLEM BY DYNAMIC TAGUCHI METHOD

Majid Jalili, Mahdi Bashiri and Amirhossein Amiri
Department of Industrial Engineering, Faculty of Engineering, Shahed University, Tehran, Iran.
Email: mjalili@shahed.ac.ir, bashiri@shahed.ac.ir, amiri@shahed.ac.ir

ABSTRACT

The Taguchi method is a useful technique to improve the performance of products or processes at a lower cost and in less time. This procedure is categorized to the static and dynamic quality characteristic.

The optimization of multiple responses has received increasing attention over the last few years in many manufacturing organizations. Several approaches dealing with multiple static quality characteristic problems have been reported. However, little attention has been focused on optimizing the multiple dynamic quality characteristics. As a result, Taguchi method is not appropriate to optimize a multi-response problem.

In this paper, we investigate multivariate Taguchi dynamic problem and propose a method based on Data Envelopment Analysis (DEA). Among the advantages of the proposed approach over traditional Taguchi method is the non-parametric, non-linear way in evaluation of all factor combinations. Simulated data shows that proposed method could increase robustness of the dynamic Taguchi method.

KEYWORDS
Data Envelopment Analysis (DEA); Taguchi Dynamic Problem; Multi Response Optimization; Signal to noise ratio; Robust parameter design.

1. INTRODUCTION

The robust design has been successfully applied to a variety of industry problems for upgrading product quality since Taguchi (1987) first introduced this method in 1980. The objective of robust design is to reduce response variation in products or processes by selecting the settings of control factors, which provide the best performance and the least sensitivity to noise factors. This act is done by using interaction between control and noise factors.

To execute the robust design, Taguchi employs an orthogonal array (OA) to arrange the experiments and uses the signal-to-noise ratio (SNR) to measure the performance of each experimental run. He proposed the dynamic SN ratio formula as follows:

\[ SN = 10 \log \left( \frac{\beta^2}{\sigma^2} \right) \]  

(1)
The $\beta$ and $\sigma$ are evaluated for different combination of control factors. SNR desires to be maximized. A two-step optimization procedure is then used to determine the optimal factor combination to simultaneously reduce the response variation and bring the mean close to the target value.

Taguchi divided the RPD methodology into two categories: static and dynamic characteristics. Static systems are defined as those desired output of the system has a fixed target value and problem attempts to obtain the value of a quality characteristic of interest as close as possible to a single specified target value. Whereas dynamic systems are those target values depends on the input signal set and there is a relationship between response (output) and signal (input). This signal-response relationship is of primary importance to the performance of the system. Figure 1(a-b) denote dynamic and static systems.

![Dynamic and Static Systems Diagram](image)

**Fig.1: (a) Dynamic system (b) Static system**

Miller and Wu (1996) criticized the terminology of Taguchi and labeled the static system, simple response system and the dynamic system, signal-response system. They said that static and dynamic systems are applied for systems, which concerned with time whereas the simple response system and signal-response systems can concern with others. They divided signal-response system into two categories, multiple target systems and measurement systems. Wu and Hamada (2000) introduced the third category as control systems.

Because the signal-response concept plays an important role in product/process development, robust parameter design of signal-response systems (also called dynamic parameter design in Taguchi’s terminology) is an effective and powerful tool for quality improvement.

Signal-response systems ideally suppose that there exist a linear relationship between the response and the signal factor. Moreover, dispersion and sensitivity are two important aspects of the signal-response system which are considered with an index such as SNR [3]. According to the linear assumption of relationship function, the simple linear regression model can be written as follows:
\[ y = \alpha + \beta M + \varepsilon \]  \hspace{1cm} (2)

where, \( M \) is the signal factor with predefined \( p \) levels with the value of \( m_i \) in \( i \)th level, \( y_i \) is the response value for \( i \)th level of signal factor, \( E(\varepsilon) = 0 \) and \( \text{Var}(\varepsilon) = \sigma^2 \). A performance measure, which converts \( \sigma^2 \) and \( \beta \) in to a single measure is

\[ \omega = \ln \left( \frac{\beta^2}{\sigma^2} \right) \]  \hspace{1cm} (3)

In contrast to Taguchi’s dynamic SNR and two-step optimization procedure, Miller and Wu [2] develop response function modeling (RFM) to optimize signal-response systems. RFM uses the experimental data to model the signal-response relationship as a function of the control and noise factors. The specified performance measure is then evaluated with respect to the fitted regression models. This method is an extension of the response modeling approach first recommended by Welch et al. (1990) and Shoemaker et al. (1991) for simple response applications.

Wasserman (1996) presents a case study of the parameter design with dynamic characteristics by using multiple regression models. Khattree (1996) provides a method for estimation in the robust parameter design in the situation in which all the noise variables cannot be studied simultaneously. He used response surface approach. Lunani et al. (1997) note that using SNR as a quality performance measure might produce inaccuracies due to a mistake to evaluate dispersion effect. They developed two graphical methods for identifying appropriate measures of dispersion, thereby avoiding interactions between the dispersion and sensitivity effects for a dynamic problem.

Miller (2002) compares three methods of analyzing signal-response applications, including Miller and Wu’s (1996) approach, Taguchi method and a graphical approach devised by Lunani et al. (1997). He introduces a new graphical technique, the joint effects plot and demonstrated usefulness of his proposed method. Lesperance and Park (2003) propose a joint generalize the linear model to evaluate the robust design of dynamic characteristics, which is based on standard regression modeling techniques.

Roshan and Wu (2002) describe the application of SNR in the analysis of the multiple target systems.

Roshan and Wu (2002) give a theoretical formulation for multiple target systems and develop a practical approach for optimization that and overcome some limitations.

Wu and Yeh (2005) present an approach to optimizing multiple dynamic problems based on quality loss. The objective is to minimize the total average quality loss for the multiple dynamic quality characteristic’s experiments. Zhiyu et al. (2006) propose a new desirability function method for multiple robust parameter design. The proposed method can yield better results than traditional desirability function approach.

Gupta et al. (2010) propose a split-plot approach to the signal-response system characterized by two variance components. They demonstrate that explicit modeling of
variance components using GLMMs leads to more precise point estimates of important model coefficients with shorter condense intervals within-profile variance and between-profile variance. Dasgupta et al. (2010) presented a robust design of measurement systems. They developed an integrated approach for estimation and reduction of measurement variation through a single parameter design experiment.


Several publications have studied the robust design problem concerning the dynamic systems such as Wasserman (1996), Lunani (1996), Miller and Wu (1996), Su and Hsieh (1998), Tsui (1997), McCaskey and Tsui (1997) and Chang (2008).

However, few studies have been concerned with optimizing the parameter design for multiple dynamic quality characteristics. Chang (2008) proposed a procedure based on desirability function to optimize multiple dynamic quality characteristics. He used Simulated Annealing to find best factor setting.

The rest of the paper is organized as follows: The proposed method is discussed in Section 2. In Section 3, a simulated dataset is applied and the performance of the proposed method is evaluated. Our concluding remarks and some future researches are given in the final section.

2. PROPOSED METHOD

DEA is a linear programming based technique for measuring the relative efficiency of a set of competing decision-making units (DMU) where comparison between inputs and outputs are difficult (Dyson et al. (1990)).

The relative efficiency of the ‘multiple inputs and outputs’ in DMU is defined as a ratio of the weighted sum of the DMU’s outputs divided by the weighted sum of the DMU’s inputs. So, if the higher performance for a DMU can be obtained, the input data of the ratio must have lower values, and the output data of the ratio must have higher values. In this article, we use DEA to select best setting of control factors in Taguchi method. Hence multiple response of signal-response system converts to the single measurement (DEA) in each level of the signal factor.

The general efficiency measure used by DEA is summarized as the following:
Max $Z_j = \frac{\sum_{r=1}^{s} u_r y_{r0}}{\sum_{i=1}^{m} v_i x_{i0}}$

\[ \sum_{r=1}^{s} u_r y_{rj} \leq 1 \]
\[ \sum_{i=1}^{m} v_i x_{ij} \]

$u_r, v_i \geq 0$

where $Z_j$: the efficiency measure of DMU $j$;

$y_r$: The values of output $y$ for DMU $j$;

$x_i$: The values of input $x$ for DMU $j$;

$u_r$: The weights assigned to trial DMU $k$ for output $y$;

$v_i$: The weights assigned to trial DMU $k$ for input $x$.

This nonlinear programming formulation (4) is equivalent to the following linear programming (LP) formulation (5) by setting its denominator equal to 1 and by maximizing its numerator.

Max $Z_j = \sum_{r=1}^{s} u_r y_{r0}$

\[ \sum_{i=1}^{m} v_i x_{ij} = 1 \]

\[ \sum_{r=1}^{s} u_r y_{rj} - \sum_{i=1}^{m} v_i x_{ij} \leq 1 \]

$u_r, v_i \geq 0$

In the basic DEA (CCR model), which developed by Charnes, Cooper, and Rhodes (1978), the objective is to maximize the relative efficiency value of a trial DMU among a reference set of DMUs; by selecting the optimal weights associated with the input and output measures. The relative efficiency value for a Taguchi dynamic problem considered as $SNR$. For a multivariate dynamic problem, $SNR$ obtained as follows:

\[ SNR = -10 \log \left[ \frac{1}{m} \sum_{k=1}^{m} y_{jk}^2 \right], \quad 0 \leq y_{jk} < \infty \]

(for the smaller-the-better response)
A data envelopment analysis method for optimizing multi-response... 

\[ \text{SNR} = -10 \log \left( \frac{1}{m} \sum_{k=1}^{m} \frac{1}{y_{ijk}^2} \right), \quad 0 \leq y_{ijk} < \infty \]  

(for the larger-the-better response) 

\[ \text{SNR} = 10 \log \left( \frac{y^2}{s^2} \right) \]  

(for the nominal-the-best response) 

where \( y_{ijk} \) is observed data for the \( i \)th response at the \( j \)th trial. 

Steps of the proposed method are expressed as: 

Step 1: Calculate \( \text{SNR} \) for each levels of signal factor. 

Step 2: Normalize the \( \text{SNR} \) for the \( i \)th response in the \( j \)th experiment. Because the \( \text{SNR} \) desire to be maximized, Equation (9) is useful. 

\[ Z_j = \frac{\text{SNR}_j - \text{SNR}_{j_{\text{min}}}}{\text{SNR}_{j_{\text{max}}} - \text{SNR}_{j_{\text{min}}}} \]  

Step 3: Estimate the relative efficiency of in each level of the signal factor for each DMU. Each treatment in the orthogonal array is regarded as a DMU when applying DEA. After that General Index (GI) is calculated as follows: 

\[ GI_i = \sqrt[12]{U_{i1} \times U_{i2} \times \ldots \times U_{i6}} \]  

where \( U_{im} \) is obtained efficiency for \( i \)th DMU in \( j \)th levels of signal factor. \( m \) is number of signal factor levels. 

Step 4: According to the calculated GI, best setting of controllable factors are selected. This setting of the control factors introduces the efficiency treatment between trials of an experiment. Note that this treatment can be an initial solution for any optimization method. 

3. NUMERICAL EXAMPLE 

We investigate the proposed index by simulated dataset. Suppose there are two responses \( \gamma_1 \) and \( \gamma_2 \). The example, including five control factors, \( x_1, x_2, x_3, x_4, x_5 \), one noise factor \( N \) and a signal factor \( M \). To consider an example, a combined array with resolution VI is used (see Table 1). Target values for each level of the signal factor for \( \gamma_1 \) are 0.1, -0.1 and 0.1 respectively. And these values for \( \gamma_2 \) are -0.2, 0.1 and 0.2.
Table 1:
Combined array $2^{6-1}_{6}$ as the designed experiment in the numerical example

<table>
<thead>
<tr>
<th>No.</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>-1</td>
<td>1</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
<td>1</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>1</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
<td>1</td>
</tr>
<tr>
<td>5</td>
<td>-1</td>
<td>-1</td>
<td>1</td>
<td>-1</td>
<td>-1</td>
<td>1</td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>-1</td>
<td>1</td>
<td>-1</td>
<td>-1</td>
<td>1</td>
</tr>
<tr>
<td>7</td>
<td>-1</td>
<td>1</td>
<td>1</td>
<td>-1</td>
<td>-1</td>
<td>1</td>
</tr>
<tr>
<td>8</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>-1</td>
<td>-1</td>
<td>1</td>
</tr>
<tr>
<td>9</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
<td>1</td>
<td>-1</td>
<td>1</td>
</tr>
<tr>
<td>10</td>
<td>1</td>
<td>-1</td>
<td>-1</td>
<td>1</td>
<td>-1</td>
<td>1</td>
</tr>
<tr>
<td>11</td>
<td>-1</td>
<td>1</td>
<td>-1</td>
<td>1</td>
<td>-1</td>
<td>1</td>
</tr>
<tr>
<td>12</td>
<td>1</td>
<td>1</td>
<td>-1</td>
<td>1</td>
<td>-1</td>
<td>1</td>
</tr>
<tr>
<td>13</td>
<td>-1</td>
<td>-1</td>
<td>1</td>
<td>1</td>
<td>-1</td>
<td>1</td>
</tr>
<tr>
<td>14</td>
<td>1</td>
<td>-1</td>
<td>1</td>
<td>1</td>
<td>-1</td>
<td>1</td>
</tr>
<tr>
<td>15</td>
<td>-1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>-1</td>
<td>1</td>
</tr>
<tr>
<td>16</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>-1</td>
<td>1</td>
</tr>
<tr>
<td>17</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>18</td>
<td>1</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>19</td>
<td>-1</td>
<td>1</td>
<td>-1</td>
<td>-1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>20</td>
<td>1</td>
<td>1</td>
<td>-1</td>
<td>-1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>21</td>
<td>-1</td>
<td>-1</td>
<td>1</td>
<td>-1</td>
<td>1</td>
<td>-1</td>
</tr>
<tr>
<td>22</td>
<td>1</td>
<td>-1</td>
<td>1</td>
<td>-1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>23</td>
<td>-1</td>
<td>1</td>
<td>1</td>
<td>-1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>24</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>-1</td>
<td>1</td>
<td>-1</td>
</tr>
<tr>
<td>25</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
<td>1</td>
<td>1</td>
<td>-1</td>
</tr>
<tr>
<td>26</td>
<td>1</td>
<td>-1</td>
<td>-1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>27</td>
<td>-1</td>
<td>1</td>
<td>-1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>28</td>
<td>1</td>
<td>1</td>
<td>-1</td>
<td>1</td>
<td>1</td>
<td>-1</td>
</tr>
<tr>
<td>29</td>
<td>-1</td>
<td>-1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>30</td>
<td>1</td>
<td>-1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>-1</td>
</tr>
<tr>
<td>31</td>
<td>-1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>-1</td>
</tr>
<tr>
<td>32</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

10 replicates are simulated in each treatment. Mean and variance-covariance matrix of two responses and obtained DEA value in each level of the signal factor (0.1, 0.2 and 0.3) are presented in Table 2.
Table 2:
Experimental results for the designed experiment of the numerical example

<table>
<thead>
<tr>
<th>No.</th>
<th></th>
<th>M1=0.1</th>
<th>M2=0.2</th>
<th>M3=0.3</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Mean (Y1)</td>
<td>Mean (Y2)</td>
<td>Var.-Cov. Matrix</td>
<td>Mean (Y1)</td>
<td>Mean (Y2)</td>
<td>Var.-Cov. Matrix</td>
</tr>
<tr>
<td>1</td>
<td>2.25</td>
<td>-1.83</td>
<td>1.12</td>
<td>0.60</td>
<td>2.87</td>
<td>-2.04</td>
<td>1.51</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.60</td>
<td>0.96</td>
<td></td>
<td></td>
<td>0.46</td>
</tr>
<tr>
<td>2</td>
<td>2.18</td>
<td>0.70</td>
<td>0.45</td>
<td>-0.12</td>
<td>1.47</td>
<td>0.97</td>
<td>0.87</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>-0.12</td>
<td>1.23</td>
<td></td>
<td></td>
<td>0.07</td>
</tr>
<tr>
<td>3</td>
<td>-1.49</td>
<td>-2.89</td>
<td>0.92</td>
<td>0.35</td>
<td>-1.38</td>
<td>-3.01</td>
<td>0.50</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.35</td>
<td>1.38</td>
<td></td>
<td></td>
<td>0.08</td>
</tr>
<tr>
<td>4</td>
<td>-3.53</td>
<td>-3.63</td>
<td>1.02</td>
<td>0.48</td>
<td>-4.18</td>
<td>-2.40</td>
<td>0.63</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.48</td>
<td>1.26</td>
<td></td>
<td></td>
<td>0.01</td>
</tr>
<tr>
<td>5</td>
<td>0.91</td>
<td>-1.68</td>
<td>1.08</td>
<td>-0.22</td>
<td>0.62</td>
<td>-2.23</td>
<td>1.23</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>-0.22</td>
<td>2.58</td>
<td></td>
<td></td>
<td>0.14</td>
</tr>
<tr>
<td>6</td>
<td>11.32</td>
<td>5.56</td>
<td>1.30</td>
<td>-0.09</td>
<td>10.77</td>
<td>5.57</td>
<td>1.15</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>-0.09</td>
<td>0.94</td>
<td></td>
<td></td>
<td>-0.44</td>
</tr>
<tr>
<td>7</td>
<td>-5.94</td>
<td>-7.26</td>
<td>0.63</td>
<td>0.11</td>
<td>-6.09</td>
<td>-7.55</td>
<td>0.86</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.11</td>
<td>1.86</td>
<td></td>
<td></td>
<td>0.27</td>
</tr>
<tr>
<td>8</td>
<td>-6.04</td>
<td>-1.54</td>
<td>1.33</td>
<td>0.53</td>
<td>-5.56</td>
<td>-0.97</td>
<td>0.96</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.53</td>
<td>1.34</td>
<td></td>
<td></td>
<td>0.07</td>
</tr>
<tr>
<td>9</td>
<td>-5.75</td>
<td>-3.40</td>
<td>0.74</td>
<td>0.46</td>
<td>-5.78</td>
<td>-3.87</td>
<td>0.37</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.46</td>
<td>1.14</td>
<td></td>
<td></td>
<td>-0.14</td>
</tr>
<tr>
<td>10</td>
<td>1.59</td>
<td>-2.85</td>
<td>0.69</td>
<td>-0.23</td>
<td>1.04</td>
<td>-2.24</td>
<td>0.44</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>-0.23</td>
<td>0.75</td>
<td></td>
<td></td>
<td>0.03</td>
</tr>
<tr>
<td>11</td>
<td>5.54</td>
<td>2.3</td>
<td>0.94</td>
<td>0.14</td>
<td>6.15</td>
<td>0.99</td>
<td>1.10</td>
</tr>
</tbody>
</table>
Table 2 (Contd.…)

<table>
<thead>
<tr>
<th>No.</th>
<th>M1=0.1</th>
<th>M2=0.2</th>
<th>M3=0.3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean (Y1)</td>
<td>Mean (Y2)</td>
<td>Mean (Y1)</td>
</tr>
<tr>
<td>12</td>
<td>-7.70</td>
<td>-2.58</td>
<td>0.45</td>
</tr>
<tr>
<td>13</td>
<td>4.69</td>
<td>7.70</td>
<td>0.63</td>
</tr>
<tr>
<td>14</td>
<td>-1.82</td>
<td>5.15</td>
<td>0.77</td>
</tr>
<tr>
<td>15</td>
<td>1.96</td>
<td>-0.55</td>
<td>1.49</td>
</tr>
<tr>
<td>16</td>
<td>4.37</td>
<td>7.95</td>
<td>0.27</td>
</tr>
<tr>
<td>17</td>
<td>7.08</td>
<td>10.63</td>
<td>1.63</td>
</tr>
<tr>
<td>18</td>
<td>-6.16</td>
<td>-5.26</td>
<td>1.57</td>
</tr>
<tr>
<td>19</td>
<td>-1.63</td>
<td>-5.00</td>
<td>0.82</td>
</tr>
<tr>
<td>20</td>
<td>4.30</td>
<td>-1.65</td>
<td>1.34</td>
</tr>
<tr>
<td>21</td>
<td>-3.59</td>
<td>2.75</td>
<td>0.42</td>
</tr>
<tr>
<td>22</td>
<td>1.69</td>
<td>-0.37</td>
<td>0.79</td>
</tr>
<tr>
<td>23</td>
<td>-4.33</td>
<td>3.28</td>
<td>0.52</td>
</tr>
</tbody>
</table>
Table 2 (Contd….)

<table>
<thead>
<tr>
<th>No.</th>
<th>M1=0.1 Mean (Y1)</th>
<th>Mean (Y2)</th>
<th>Var.-Cov. Matrix</th>
<th>M2=0.2 Mean (Y1)</th>
<th>Mean (Y2)</th>
<th>Var.-Cov. Matrix</th>
<th>M3=0.3 Mean (Y1)</th>
<th>Mean (Y2)</th>
<th>Var.-Cov. Matrix</th>
<th>M1=0.1</th>
<th>M2=0.2</th>
<th>M3=0.3</th>
</tr>
</thead>
<tbody>
<tr>
<td>24</td>
<td>9.63</td>
<td>5.10</td>
<td></td>
<td>10.74</td>
<td>5.16</td>
<td></td>
<td>11.05</td>
<td>4.37</td>
<td></td>
<td>0.17</td>
<td>0.22</td>
<td>0.37</td>
</tr>
<tr>
<td></td>
<td>0.99</td>
<td>0.19</td>
<td></td>
<td>0.84</td>
<td>0.26</td>
<td></td>
<td>0.65</td>
<td>0.23</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.19</td>
<td>1.21</td>
<td></td>
<td>0.26</td>
<td>0.51</td>
<td></td>
<td>0.23</td>
<td>0.91</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>4.78</td>
<td>2.25</td>
<td></td>
<td>4.32</td>
<td>2.20</td>
<td></td>
<td>3.70</td>
<td>2.13</td>
<td></td>
<td>0.36</td>
<td>0.12</td>
<td>0.28</td>
</tr>
<tr>
<td></td>
<td>1.56</td>
<td>-0.08</td>
<td></td>
<td>0.32</td>
<td>-0.04</td>
<td></td>
<td>1.59</td>
<td>1.02</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>-0.08</td>
<td>0.66</td>
<td></td>
<td>-0.04</td>
<td>0.42</td>
<td></td>
<td>1.02</td>
<td>1.59</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>26</td>
<td>-6.64</td>
<td>-0.33</td>
<td></td>
<td>-7.77</td>
<td>-0.39</td>
<td></td>
<td>-8.56</td>
<td>0.00</td>
<td></td>
<td>0.47</td>
<td>0.68</td>
<td>0.79</td>
</tr>
<tr>
<td></td>
<td>0.33</td>
<td>0.01</td>
<td></td>
<td>1.32</td>
<td>-0.28</td>
<td></td>
<td>0.65</td>
<td>-0.09</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.01</td>
<td>0.83</td>
<td></td>
<td>1.13</td>
<td></td>
<td></td>
<td>-0.09</td>
<td>0.43</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>27</td>
<td>-0.78</td>
<td>-5.48</td>
<td></td>
<td>-0.06</td>
<td>-5.39</td>
<td></td>
<td>0.58</td>
<td>-5.03</td>
<td></td>
<td>0.97</td>
<td>0.93</td>
<td>0.91</td>
</tr>
<tr>
<td></td>
<td>1.12</td>
<td>-0.23</td>
<td></td>
<td>1.93</td>
<td>0.09</td>
<td></td>
<td>1.04</td>
<td>-0.02</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>-0.23</td>
<td>0.82</td>
<td></td>
<td>0.09</td>
<td>1.13</td>
<td></td>
<td>-0.02</td>
<td>1.73</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>28</td>
<td>-10.03</td>
<td>-6.39</td>
<td></td>
<td>-10.53</td>
<td>-7.75</td>
<td></td>
<td>-10.77</td>
<td>-7.70</td>
<td></td>
<td>0.16</td>
<td>0.18</td>
<td>0.12</td>
</tr>
<tr>
<td></td>
<td>0.63</td>
<td>0.24</td>
<td></td>
<td>0.69</td>
<td>0.12</td>
<td></td>
<td>1.90</td>
<td>0.05</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.24</td>
<td>1.05</td>
<td></td>
<td>0.12</td>
<td>0.43</td>
<td></td>
<td>0.05</td>
<td>2.08</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>29</td>
<td>-2.06</td>
<td>-1.63</td>
<td></td>
<td>-1.89</td>
<td>-1.58</td>
<td></td>
<td>-2.28</td>
<td>-1.30</td>
<td></td>
<td>0.14</td>
<td>0.16</td>
<td>0.29</td>
</tr>
<tr>
<td></td>
<td>0.77</td>
<td>0.50</td>
<td></td>
<td>0.79</td>
<td>-0.09</td>
<td></td>
<td>0.59</td>
<td>0.41</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.50</td>
<td>2.00</td>
<td></td>
<td>-0.09</td>
<td>0.33</td>
<td></td>
<td>0.41</td>
<td>1.54</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>1.72</td>
<td>4.05</td>
<td></td>
<td>2.29</td>
<td>3.68</td>
<td></td>
<td>2.77</td>
<td>3.42</td>
<td></td>
<td>0.56</td>
<td>0.42</td>
<td>0.34</td>
</tr>
<tr>
<td></td>
<td>1.22</td>
<td>0.56</td>
<td></td>
<td>0.81</td>
<td>0.53</td>
<td></td>
<td>0.83</td>
<td>0.95</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.56</td>
<td>0.60</td>
<td></td>
<td>0.53</td>
<td>1.98</td>
<td></td>
<td>0.95</td>
<td>3.14</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>31</td>
<td>-3.76</td>
<td>-4.96</td>
<td></td>
<td>-3.87</td>
<td>-5.01</td>
<td></td>
<td>-4.83</td>
<td>-5.00</td>
<td></td>
<td>0.89</td>
<td>0.95</td>
<td>0.88</td>
</tr>
<tr>
<td></td>
<td>0.50</td>
<td>-0.16</td>
<td></td>
<td>0.15</td>
<td>0.18</td>
<td></td>
<td>2.30</td>
<td>0.48</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>-0.16</td>
<td>0.85</td>
<td></td>
<td>0.18</td>
<td>0.64</td>
<td></td>
<td>0.48</td>
<td>1.14</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>32</td>
<td>9.19</td>
<td>3.22</td>
<td></td>
<td>9.28</td>
<td>3.44</td>
<td></td>
<td>10.44</td>
<td>3.31</td>
<td></td>
<td>0.34</td>
<td>0.53</td>
<td>0.19</td>
</tr>
<tr>
<td></td>
<td>1.22</td>
<td>-0.03</td>
<td></td>
<td>0.87</td>
<td>0.46</td>
<td></td>
<td>1.00</td>
<td>-0.04</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>-0.03</td>
<td>0.95</td>
<td></td>
<td>0.46</td>
<td>1.71</td>
<td></td>
<td>-0.04</td>
<td>1.16</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
To select best treatment, a single measurement (GI) is calculated for each treatment. Table 3 shows results of obtained GI for each treatment. Note that GI is geometric mean of calculated DEA in signal factor levels.

**Table 3: Computed general index for each treatment in the numerical example**

<table>
<thead>
<tr>
<th>Treatment</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>GI</td>
<td>0.207</td>
<td>0.445</td>
<td>0.25</td>
<td>0.293</td>
<td>0.153</td>
<td>0.502</td>
<td>0.903</td>
<td>0.185</td>
</tr>
<tr>
<td>Treatment</td>
<td>9</td>
<td>10</td>
<td>11</td>
<td>12</td>
<td>13</td>
<td>14</td>
<td>15</td>
<td>16</td>
</tr>
<tr>
<td>GI</td>
<td>0.170</td>
<td>0.298</td>
<td>0.791</td>
<td>0.403</td>
<td>0.233</td>
<td>0.252</td>
<td>0.933</td>
<td>0.166</td>
</tr>
<tr>
<td>Treatment</td>
<td>17</td>
<td>18</td>
<td>19</td>
<td>20</td>
<td>21</td>
<td>22</td>
<td>23</td>
<td>24</td>
</tr>
<tr>
<td>GI</td>
<td>0.309</td>
<td>0.808</td>
<td>0.836</td>
<td>0.160</td>
<td>0.642</td>
<td>0.450</td>
<td>0.542</td>
<td>0.240</td>
</tr>
<tr>
<td>Treatment</td>
<td>25</td>
<td>26</td>
<td>27</td>
<td>28</td>
<td>29</td>
<td>30</td>
<td>31</td>
<td>32</td>
</tr>
<tr>
<td>GI</td>
<td>0.227</td>
<td>0.632</td>
<td>0.936</td>
<td>0.151</td>
<td>0.187</td>
<td>0.431</td>
<td>0.906</td>
<td>0.325</td>
</tr>
</tbody>
</table>

According to the proposed GI, treatment 27 with maximum value of GI (0.936) and 31 with 0.906 value of GI are selected as proper treatments for experiment. Figure 2 displays responses $Y_1$ and $Y_2$ according to the signal factor.

![Figure 2](a-1) ![Figure 2](a-2)
A data envelopment analysis method for optimizing multi-response...

Fig. 2: Shape of responses
(a-1) Treatment 31, (a-2) Treatment 27
(a-3) Treatment 5,  (a-4) Treatment 22

Figure 2 depicted treatments 31 and 27 as proposed treatment versus treatments 5 and 22 as random treatments from the designed experiment. As Figure 2 (a-1) and (a-2) shows, proposed setting for control factors make responses robust and there is less variation in the responses in levels of the signal factor. In contrast, responses for treatments 5 and 22 (see Figure 2 (a-3) and (a-4)) have large variation and there is a large distance between responses and their targets.

4. CONCLUSION

In this paper, a method based on Data Envelopment Analysis (DEA) is proposed to consider multivariate signal-response systems. By using DEA, location and dispersion of responses can be considered with a single measurement (GI) in each level of the signal factor. Finally, proper control factor levels are obtained by considering GI in each level of the signal factor. The simulated result shows efficiency of the proposed method. To find the optimal value of controllable factors in a continuous space, modeling of GI according to the control factors as general linear model is recommended as a future research.

REFERENCES

A NOTE ON BATHTUB HAZARD RATE

Sajeela Tazeen Dara\(^1\) and Munir Ahmad\(^2\)
\(^1\) Government Islamia College for Women, Lahore, Pakistan
Email: sjeelaqm@gmail.com
\(^2\) National College of Business Administration & Economics, Lahore, Pakistan. Email: drmunir@brain.net.pk

ABSTRACT

The hazard rate function has a pivotal role in actuarial sciences, reliability and survival analysis. A simple random sample cannot be taken from entire population in these areas of study. In this paper, we examine the hazard rates of some moment distributions, particularly bathtub hazard rates. The bathtub hazard rate represents failure behavior of items / individuals in reliability engineering and survival analysis.

1. INTRODUCTION


The hazard rate function (HRF) measures the conditional instantaneous rate of failure at time \(x\), given survival to time \(x\). This is known as the force of mortality in actuarial sciences and demography, as intensity rate in extreme value theory and as inverse Mill’s Ratio in duration analysis in economics and Sociology. The hazard rate function plays the most important role in reliability and survival analysis. This is also named as instantaneous failure rate.

The hazard rate function, regardless of its terminology, is defined as

\[
h(x) = \frac{f(x)}{1 - F(x)} = \frac{f(x)}{S(x)},
\]

where \(f(x)\): the probability density function, \(F(x)\): the probability density function and \(S(x)\): the survival function, also denoted as \(\bar{F}(x)\).
The hazard rate function may loosely classify into three groups:

i) Monotonic hazard rates: \( h(x) \) is either increasing or decreasing.

ii) Bathtub hazard rates: \( h(x) \) is bathtub or U-shaped.

iii) Generalized bathtub hazard rates: \( h(x) \) is a polynomial.


2. MOMENT DISTRIBUTIONS

In statistical analysis, a particular characteristic, say \( X \), is always under consideration. A random sample \( (X_1, X_2, \ldots, X_n) \) is taken from the population; the distribution of observed \( X_i \)'s is not the same as that of the original distribution of \( X \).

Suppose the random variable \( X \) has a probability distribution \( f(x; \theta) \), with unknown parameter \( \theta \). The corresponding distribution, called moment distribution, is of the form

\[
g(x; \theta) = \frac{m(x) f(x; \theta)}{E[m(X)]},
\]

where \( m(x) \) is a non-negative weight function such that \( E[m(X)] \) exists.

When the weight function has the form, \( m(x) = x^\alpha \) then such distributions are named as size-biased distributions of order \( \alpha \) and are written as [Patil and Ord (1976); Patil (1981) and Mahfoud and Patil (1982)]:

\[
g(x; \theta) = \frac{x^\alpha f(x; \theta)}{\mu'_\alpha}
\]
where $\mu'_\alpha = \int x^\alpha f(x; \theta) \, dx$ is the $\alpha^{th}$ raw moment of $f(x; \theta)$.

When $\alpha = 1$ or $2$, these special cases are termed as length-biased or size-biased distribution and area-biased distribution, respectively.

Equilibrium distribution and residual life distribution may also be derived from moment distribution with inverse of the hazard rate function and $I_{[x>t]}$ function, as the weight function, respectively.

### 3. BATHTUB HAZARD RATE

The bathtub hazard rate (BHR) represents failure behavior of items / individuals in survival analysis, reliability analysis, duration analysis etc. This shows the variation of the hazard rate function or failures of items / individuals during their lifetime. The bathtub hazard rate depicts the relative hazard rate of the entire population over time, not hazard rate of a single time. This is an example of treating more than one failure type by a single classification.

The bathtub hazard rate curve comprises of three time periods; viz. (i) Burn-in, (ii) Useful-life and (iii) Wear-out. The hazard rate function $h(x)$ decreases with time at first (Burn-in), then remains constant with respect to time (useful-life) and it increases with time (wear-out). The bathtub is modeled by a piecewise set of three hazard rate functions.

Suppose the random variable $X$ has standardized beta probability distribution

$$f(x) = \frac{x^{p-1}(1-x)^{q-1}}{B(p,q)}, 0 \leq x \leq 1, p, q > 0,$$

then the corresponding moment distribution has probability density function

$$g(x) = \frac{x^{p+\alpha-1}(1-x)^{q-1}}{B(q, p + \alpha)}, 0 < x < 1, p, q > 0, \alpha = 1, 2$$

and the hazard rate function is

$$h_g(x) = \frac{x^{p-1}(1-x)^{q-1}}{B(p,q) - B(x; p, q)}.$$

The graphs of moment beta pdf and its corresponding bath-tub hazard rate are
Some Moment distributions, modeling bathtub shaped hazard rate function, are summarized in Table 1.

<table>
<thead>
<tr>
<th>Moment Distribution</th>
<th>Prob. Density Function</th>
<th>Hazard Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chi Squared</td>
<td>[ g(x) = \frac{x^\alpha f(x)}{\int_0^\infty x^\alpha f(x) dx} ]</td>
<td>[ h_g(x) = \frac{g(x)}{1-G(x)} ]</td>
</tr>
<tr>
<td>( \alpha = 1 )</td>
<td>[ \frac{1}{2^{\frac{\alpha}{2}} \Gamma \left( \frac{\alpha}{2} \right)} x^{\frac{\alpha}{2} - 1} \exp \left( -\frac{x}{2} \right) ]</td>
<td>[ x^{\frac{\alpha}{2} + 1} \exp \left( -\frac{x}{2} \right) ]</td>
</tr>
<tr>
<td>( \alpha = 2 )</td>
<td>[ 3.000E+00 ]</td>
<td>[ 5.000E-01 ]</td>
</tr>
<tr>
<td>Distribution</td>
<td>Moment Cumulative Distribution Function</td>
<td>Moment Density Function</td>
</tr>
<tr>
<td>-------------------------------------</td>
<td>-----------------------------------------</td>
<td>-------------------------</td>
</tr>
<tr>
<td>Extreme Value-Minimum</td>
<td>$x^\alpha \exp \left( \frac{x-\mu}{\beta} \right) \exp \left( -\exp \left( \frac{x-\mu}{\beta} \right) \right)$</td>
<td>$\int_0^\infty x^\alpha \exp \left( \frac{x-\mu}{\beta} \right) \exp \left( -\exp \left( \frac{x-\mu}{\beta} \right) \right) dx$</td>
</tr>
<tr>
<td>Pareto</td>
<td>$\frac{x^{\alpha-a-1} (a-\alpha)}{b^{\alpha-a}}$</td>
<td>$b^{\alpha-a}$</td>
</tr>
<tr>
<td></td>
<td>$b &lt; x &lt; \infty, a &gt; 0, b &gt; 0$</td>
<td>$\alpha = 1$</td>
</tr>
<tr>
<td></td>
<td>$\alpha = 1$</td>
<td>$\alpha = 2$</td>
</tr>
<tr>
<td>Rectangular</td>
<td>$x^\alpha (1+\alpha), 0 &lt; x &lt; 1$</td>
<td>$\frac{x^\alpha (1+\alpha)}{1-x^{1+\alpha}}$</td>
</tr>
</tbody>
</table>
REFERENCES

MIXTURE RATIO AND REGRESSION ESTIMATORS USING MULTI-AUXILIARY VARIABLES AND ATTRIBUTES IN SINGLE PHASE SAMPLING

Muhammad Moeen¹, Muhammad Qaiser Shahbaz² and Muhammad Hanif¹
¹ National College of Business Administration & Economics, Lahore, Pakistan. Email: mmoeenbutt65@gmail.com and hanif@lums.edu.pk
² Department of Mathematics COMSATS Institute of Information Technology, Lahore, Pakistan. Email: qshahbaz@gmail.com

ABSTRACT
In this paper a class of mixture ratio and regression estimators for single phase sampling have been proposed by using information on auxiliary variables and attributes simultaneously. Mean square errors of the proposed estimators have also been derived and a real life example has been taken to compare their efficiency.

KEYWORD
Mixture ratio estimators; mixture regression estimators.

1. INTRODUCTION
There are situations in which while estimating the mean of the study variable one can improve the estimation procedure by making use of available auxiliary information on both variables and attributes which are strongly associated with the variable of interest. For example if we are interested to estimate the mean yield of rice per acre where the auxiliary variables and attributes associated with the yield of rice include fertility of land, water availability, type of seed, amount of rainfall, type and quantity of fertilizer etc. These variables and attributes can be incorporated with the main variable to form ratio, regression or product estimators of the population mean. Various researchers in the past integrated auxiliary variables and attributes independently with the study variable to form efficient estimators. Cochran (1940) proposed a ratio estimator using one auxiliary variable under single phase sampling. Hansen et al. (1953) proposed a regression estimator to estimate population mean by using one auxiliary variable in single phase sampling. Samiuddin and Hanif (2006) and Ahmad (2008) proposed ratio and regression estimators using multi auxiliary variables in single-phase sampling. Singh et al. (2008) used an auxiliary attribute and proposed a ratio estimator in single phase sampling. Hanif et al. (2009) and Hanif et al. (2010) proposed family of regression and ratio type estimators using multi auxiliary attributes. In this paper we combine auxiliary variables and attributes with the study variable and propose mixture ratio and regression estimators in single phase sampling.

2. NOTATIONS
Let a simple random sample of size n units without replacement is to be drawn from a population of N units to estimate the population mean of the variable Y. Let there be ‘n’
auxiliary variables $X_1, X_2, X_3, ..., X_k, ..., X_n$ available and the let the population may be completely dichotomized with respect to presence or absence of an attribute $\tau_j$, $j = 1, 2, ..., m$, assuming that the attribute $\tau_{ij}$ takes only values ‘0’ or ‘1’ accordingly as:

$$\tau_{ij} = 1 \text{ if the } i^{th} \text{ unit of the population possesses attribute } \tau_j,$$
$$= 0 \text{ otherwise.}$$

Let for the ‘m’ auxiliary attributes $\tau_1, \tau_2, \tau_3, ..., \tau_m$,

$$A_j = \sum_{i=1}^{N} \tau_{ij} : \text{ total number of units in the population possessing attribute } \tau_j,$$

$$a_j = \sum_{i=1}^{n} \tau_{ij} : \text{ total number of units in the sample possessing attribute } \tau_j,$$

$$P_j = \frac{A_j}{N} : \text{ proportion of units in the population possessing attribute } \tau_j,$$

$$p_j = \frac{a_j}{n} : \text{ proportion of units in the sample possessing attribute } \tau_j,$$

Let

$$\theta = \frac{1}{n} - \frac{1}{N}, \quad \bar{y} = \bar{Y} + \bar{e}_y, \quad \bar{x}_i = \bar{X} + \bar{e}_{x_i}, \quad p_j = P_j + \bar{e}_{\tau_j}, \quad E\left(\bar{e}_y^2\right) = \theta \bar{Y}^2 C_y,$$

$$E\left(\bar{x}_i^2\right) = \theta \bar{X}^2 C_x, \quad E\left(\bar{e}_{\tau_j}^2\right) = \theta P_j^2 C_{\tau_j}, \quad E\left(\bar{e}_y \bar{e}_{\tau_j}\right) = \theta \bar{Y} P_j C_y C_{\tau_j} \rho_{pb_j},$$

$$E(\bar{e}_x \bar{e}_{\tau_j}) = \theta \bar{X} P_j C_x C_{\tau_j} \rho_{pb_j}, \quad S_{y\tau_j} = \frac{1}{N-1} \sum_{i=1}^{N} (y_i - \bar{Y})(\tau_{ij} - P_j),$$

point bi-serial correlation coefficient between auxiliary attribute and study variable

$$\rho_{p_{x_k} b_j} = S_{x_k \tau_j} / \left(S_{x_k} S_{\tau_j}\right) : \text{ point bi-serial correlation coefficient between auxiliary attribute and variable } Q_{js} (-1 \leq Q_{js} \leq 1) : \text{ Phi coefficient between attributes } \tau_j \text{ and } \tau_s,$$

$$\Delta_{(x_x)} : \text{ matrix of point bi-serial correlations between one auxiliary variable and m-auxiliary attributes and phi-coefficients among } m \text{ attributes of order } m+1 = q,$$

$$\Delta_{(x_{\tau_1})} : \text{ matrix of point bi-serial correlations between } n \text{ auxiliary variables and one auxiliary attribute and correlations among } n \text{ auxiliary variables of order } n+1 = r,$$

$$\left(\Delta_{y_{x_k}}^{*}\right)_{x, \tau_i} : \text{ matrix of point bi-serial correlations of order } n+1 = r \text{ for } k = 1, 2, ..., n$$
\[
\left( \Delta^*_{y_{ij}} \right)_{i,x_1} : \text{ matrix of point bi.Serial correlations and phi coefficients of order} \\
q = m+1 \text{ for } j = 1, 2, \ldots, m
\]

\[a_{q1} : q \times 1 \text{ vector of } \alpha \text{ coefficients; } [(m+1) = q] \]

\[b_{r1} : r \times 1 \text{ vector of } \beta \text{ coefficients; } [(n+1) = r] \]

\[c_{r1} : r \times 1 \text{ vector of correlation coefficients and a point bi.Serial correlation coefficient} \]

\[d_{q1} : q \times 1 \text{ vector of point bi.Serial correlation coefficients and a correlation coefficient} \]

3. SOME PREVIOUS ESTIMATORS IN SINGLE PHASE SAMPLING

In the discussion that follows we present some well known estimators and their mean square errors in single phase sampling using auxiliary variables and attributes.

Laplace (1786) ratio estimator:

\[T_{l1} = \frac{\bar{y}}{\bar{x}} \frac{\bar{X}}{\bar{X}}, \quad \text{MSE}\left( T_{l1} \right) = \theta_2 \bar{Y}^2 \left[ C_y^2 + C_x^2 - 2 \rho_{xy} C_x C_y \right]. \]

Cochran (1940) regression estimator:

\[T_{21} = \left( \bar{y} + b_{yx} \left( \bar{X} - \bar{x} \right) \right), \quad \text{MSE}\left( T_{21} \right) = \theta \bar{Y}^2 C_y^2 \left[ 1 - \rho_{xy}^2 \right]. \]

Srivastava (1967) ratio estimator:

\[T_{31} = \frac{\bar{y}}{\bar{x}} \left( \frac{\bar{X}}{\bar{x}} \right)^a, \quad \text{MSE}\left( T_{31} \right) = \theta \bar{Y}^2 C_y^2 \left[ 1 - \rho_{xy}^2 \right]. \]

Samiuddin and Hanif (2006) ratio estimator:

\[T_{41} = \frac{\bar{y}}{\bar{x}} \frac{\bar{Z}}{\bar{Z}}, \quad \text{MSE}\left[ T_{41} \right] = \theta \bar{Y}^2 \left[ C_x^2 + C_y^2 + C_z^2 - 2 C_x C_y \rho_{xy} - 2 C_y C_z \rho_{yz} + 2 C_x C_z \rho_{xz} \right]. \]

Naik and Gupta (1996) ratio estimator:

\[T_{51} = \frac{\bar{y}}{\bar{p}} \frac{p_1}{p_t}, \quad \text{MSE}\left( T_{51} \right) = \theta \bar{Y}^2 \left[ C_y^2 + C_{\tau_1}^2 - 2 \rho_{ph} C_y C_{\tau_1} \right]. \]

Naik and Gupta (1996) regression estimator:

\[T_{61} = \bar{y} + \left( p_1 - P_1 \right) \rho, \quad \text{MSE}\left( T_{61} \right) = \theta \left( 1 - \rho_{ph}^2 \right) \bar{Y}^2 C_y^2 \]
Shabbir and Gupta (2007) estimator:
\[ t_{7(1)} = \left[ d_1 \bar{y} + d_2 \left( P_1 - p_1 \right) \right] \frac{P_1}{p_1}, \text{ for } p_1 > 0 \]
\[
MSE \left( t_{7(1)} \right) \approx \frac{\theta \left( 1 - \rho_{\theta \rho_0} \right) \bar{Y}^2 C_y^2}{1 + \theta \left( 1 - \rho_{\theta \rho_0} \right) C_y^2}.
\]

Haq (2010) generalized ratio estimator:
\[ t_{8(1)} = \bar{y} \left( \frac{P_1}{p_1} \left( \frac{P_2}{p_2} \right) \cdots \left( \frac{P_k}{p_k} \right) \right), \]
\[
MSE \left( t_{8(1)} \right) \approx \bar{Y}^2 C_y^2 \left[ \theta \left( 1 - \rho_{\bar{y}, \tau_1, \tau_2, \ldots, \tau_k} \right) \right].
\]

4. THE PROPOSED ESTIMATORS

MIXTURE Ratio Estimators in Single phase sampling

**Theorem 4.1**

Let \( \phi^*_y(z, x_i) \) be a mixture ratio estimator based on single auxiliary variable and ‘m’ auxiliary attributes given by:

\[
\phi^*_y(z, x_i) = \bar{y} \left( \frac{X_1}{X_1} \right)^{\beta_1} \prod_{j=1}^{m} \left( \frac{P_j}{p_j} \right)^{\alpha_j}.
\]

The mean square error of \( \phi^*_y(z, x_i) \) is

\[
MSE \left( \phi^*_y(z, x_i) \right) = \frac{\theta \bar{Y}^2 C_y^2}{\left| \Delta_{q \times q} \right|} \left| \Delta_{(y, x_j)} \right|.
\]

**Proof:**

We may write

\[ \phi^*_y(z, x_i) = \bar{y} \left( \frac{X_1}{X_1} \right)^{\beta_1} \prod_{j=1}^{m} \left( \frac{P_j}{p_j} \right)^{\alpha_j} \] \hspace{1cm} (4.1.1)

We may write (4.1.1) by using the usual notations (section 2.) as,

\[
\phi^*_y(z, x_i) = \left( \bar{y} + \bar{e}_y \right) \left( 1 - \frac{\beta \bar{e}_x}{X_1} \right) \prod_{j=1}^{m} \left( 1 - \frac{\alpha_j \bar{e}_{x_j}}{P_j} \right)
\]
or
\[
\phi^*_y(z, x_i) = \bar{y} + \bar{e}_y - \frac{\bar{y}}{X_1} \beta \bar{e}_x - \sum_{j=1}^{m} \frac{\bar{y}}{P_j} \alpha_j \bar{e}_{x_j},
\]
or
$$E \left( \phi^*_{y(x_i,x_j)} - \bar{Y} \right)^2 = E \left[ \bar{e}_y - \frac{\bar{Y}}{X_1} \beta \bar{e}_{x_j} - \sum_{j=1}^{m} \frac{\bar{Y}}{P_j} \alpha_j \bar{e}_{x_j} \right]^2$$

(4.1.2)

Partially differentiating (4.1.2) w.r.t. $\beta$ and $\alpha_j$, ($j=1,2,\ldots,m$) and equating the results to zero we obtain the optimum values of $\beta$ and $\alpha_j$’s by solving the ‘$m+1=q$’ Normal equations given by,

$$\Delta_{(x_i,x_j)} b_{r\times 1} = C_y d_{q\times 1}$$

or

$$b = C_y \Delta_{(x_i,x_j)}^{-1} d,$$

provided $|\Delta| \neq 0$

$$b = C_y \frac{\text{Adj} \left( \Delta_{(x_i,x_j)} \right) d}{|\Delta_{(x_i,x_j)}|}.$$

From (4.1.2) we get,

$$MSE \left( \phi^*_{y(x_i,x_j)} \right) = \theta \bar{Y}^2 \left[ C_y^2 - \beta C_x C_y \rho_{x,y} - \sum_{j=1}^{m} \alpha_j C_x C_t \rho_{P,b_j} \right]$$

(4.1.3)

Substituting the values of $\alpha_j$’s and $\beta$ in (4.1.3) we get,

$$MSE \left( \phi^*_{y(x_i,x_j)} \right) = \theta \bar{Y}^2 C_y^2 \left[ 1 + \frac{(-1)^{q+1} |\Delta_{y,x_i}| \Delta_{x_i,x_j} C_x \rho_{x,y} + \sum_{j=1}^{m} \frac{(-1)^{i+1} |\Delta_{y,t_j}| \Delta_{x_i,x_j} C_t \rho_{P,b_j}}{|\Delta_{(x_i,x_j)}|} \right]$$

or

$$MSE \left( \phi^*_{y(x_i,x_j)} \right) = \frac{\theta \bar{Y}^2 C_y^2}{|\Delta_{(x_i,x_j)}|} \left[ |\Delta_{y,x_i,x_j}| \right]$$

(4.1.4)

**Corollary 4.1**

The mean square error of $\phi^*_{y(x_i,x_j)} = \bar{y} \left( \frac{P_i}{p_1} \right)^{\alpha_i} \left( \frac{X_1}{x_i} \right)^{\beta_i}$ is given by,

$$MSE \left( \phi^*_{y(x_i,x_j)} \right) = \frac{\theta \bar{Y}^2 C_y^2}{|\Delta_{(x_i,x_j)}|} \left[ |\Delta_{y,x_i,t_j}| \right].$$

where

$$|\Delta_{(y,x_i,x_j)}| = \begin{vmatrix} 1 & \rho_{x,y} & \rho_{P,b_i} \\ \rho_{x,y} & 1 & \rho_{P,b_h} \\ \rho_{P,b_h} & \rho_{P,b_h} & 1 \\ \end{vmatrix}, \quad |\Delta_{(x_i,t_j)}| = \begin{vmatrix} 1 & \rho_{P,b_i} \\ \rho_{P,b_h} & 1 \\ \end{vmatrix}.$$
Corollary 4.2

The mean square error of \( \hat{\phi}_{(\tau_1, \tau_2, x_i)}^* \) = \( \bar{y} \left( \frac{P_1}{P_1} \right)^{\alpha_1} \left( \frac{P_2}{P_2} \right)^{\alpha_2} \left( \frac{\bar{x}_1}{\bar{x}_1} \right)^{\beta_1} \) is given by,

\[
MSE\left( \hat{\phi}_{(\tau_1, \tau_2, x_i)}^* \right) = \frac{\theta \bar{y}^2 C_y^2}{|\Delta_{(\tau_1, \tau_2, x_i)}|} \left[ |\Delta_{(y, x_1, \tau_2, \tau_2)}| \right].
\]

where

\[
|\Delta_{(y, x_1, \tau_1, \tau_2)}| = \begin{vmatrix}
1 & \rho_{x_1y} & \rho_{P_1b_1} & \rho_{P_2b_2} \\
\rho_{x_1y} & 1 & \rho_{P_1b_1} & \rho_{P_2b_2} \\
\rho_{P_1b_1} & \rho_{P_1b_1} & 1 & Q_{12} \\
\rho_{P_2b_2} & \rho_{P_2b_2} & Q_{12} & 1 \\
\end{vmatrix},
\]

\[
|\Delta_{(\tau_1, \tau_2, y)}| = \begin{vmatrix}
1 & \rho_{P_1b_1} & \rho_{P_2b_2} \\
\rho_{P_1b_1} & 1 & Q_{12} \\
\rho_{P_2b_2} & Q_{12} & 1 \\
\end{vmatrix}.
\]

5. Mixture Regression Estimators
In Single Phase Sampling

Theorem 5.1

Let \( \gamma_{(x,y)} \) be a mixture regression estimator based on ‘n’ auxiliary variables and one auxiliary attribute given by,

\[
\gamma_{(x,y)} = \bar{y} + \alpha_1^* (P_1 - p_1) + \sum_{i=1}^{n} \beta_i^* (\bar{x}_i - \bar{x}).
\]

The mean square error of \( \gamma_{(x,y)} \) is

\[
MSE\left( \gamma_{(x,y)} \right) = \frac{\theta \bar{y}^2 C_y^2}{|\Delta_{(x,y)}|} \left[ |\Delta_{(y,x,y)}| \right].
\]

Proof:

We may write

\[
\gamma_{(x,y)} = \bar{y} + \alpha_1^* (P_1 - p_1) + \sum_{i=1}^{n} \beta_i^* (\bar{x}_i - \bar{x}).
\]

using the given results as,

\[
\gamma_{(x,y)} - \bar{Y} = \bar{e}_y - \alpha_1^* \bar{e}_1 - \sum_{i=1}^{n} \beta_i^* \bar{e}_x_i,
\]

or

\[
MSE\left( \gamma_{(x,y)} \right) = E \left[ \bar{e}_y - \alpha_1^* \bar{e}_1 - \sum_{i=1}^{n} \beta_i^* \bar{e}_x_i \right]^2 \tag{4.1.2}
\]

Partially differentiating (4.1.2) w.r.t. \( \alpha_1 \) and \( \beta_i \)’s \( (i = 1, 2, \ldots, n) \) and equating the results to zero we obtain the optimum values of \( \alpha_1 \) and \( \beta_i \)’s by solving the ‘ \( n + 1 = q \) ’ Normal equations given by,
\[
\Delta^*_q b^*_{q+1} = \bar{Y}C_y d_{q+1},
\]
or
\[
b^* = \bar{Y}C_y \Delta^*_q d,
\]
or
\[
b^* = \bar{Y}C_y \frac{\text{Adj}(\Delta^*_q) d}{|\Delta^*_q|}
\]
provided \( |\Delta^*_q| \neq 0 \).

(4.1.3)

or,
\[
\begin{bmatrix}
\bar{X}_1 C_{x_1} \beta_1 \\
\bar{X}_2 C_{x_2} \beta_2 \\
\vdots \\
\bar{X}_n C_{x_n} \beta_n \\
P C_t \alpha_1
\end{bmatrix}
\begin{bmatrix}
(-1)^{i+1} |\Delta_{x_i x_i}|_{\xi, \tau_i} \\
(-1)^{2+i+1} |\Delta_{y,x_j}|_{\xi, \tau_i} \\
\vdots \\
(-1)^{q+1} |\Delta_{y,y_j}|_{\xi, \tau_i}
\end{bmatrix}
= \bar{Y}C_y \Delta^*_q M
\]

(4.1.4)

where \( q = n + 1 \).

Rewriting (4.1.2) we get,
\[
\text{MSE} \left( \gamma_{y(\xi, \tau_i)} \right) = E \left( \bar{e} \right) \left[ \bar{e} - \alpha_i \bar{e}_{x_i} - \sum_{i=1}^{n} \beta_i \bar{e}_{x_i} \right]
\]

(4.1.5)

Using the notations in (4.1.5) given in section 2. we get,
\[
\text{MSE} \left( \gamma_{y(\xi, \tau_i)} \right) = \theta \left[ \bar{Y}^2 C_y^2 - \alpha_i \bar{Y}C_y C_t \rho_{P_y b} - \bar{Y}C_y \sum_{i=1}^{n} \beta_i \bar{X}_i C_{x_i} \rho_{x_i y} \right]
\]

(4.1.6)

Substituting the values of \( \alpha_1 \) and \( \beta_i \)'s in (4.1.6) we get,
\[
\text{MSE} \left( \gamma_{y(\xi, \tau_i)} \right) = \theta \bar{Y}^2 C_y^2 \frac{(-1)^{i+1} |\Delta_{y,x_i}|_{\xi, \tau_i}}{|\Delta^*_q|} P_t C_t \rho_{P_y b_t}
\]

\[
+ \theta \bar{Y}^2 C_y^2 \sum_{i=1}^{n} \frac{(-1)^{i+1} |\Delta_{y,x_i}|_{\xi, \tau_i}}{|\Delta^*_q|} \bar{X}_i C_{x_i} \rho_{x_i y},
\]
or
\[
\text{MSE} \left( \gamma_{y(\xi, \tau_i)} \right) = \frac{\theta \bar{Y}^2 C_y^2}{|\Delta^*_q|} \left[ |\Delta^*_q| + (-1)^{q+1} \sum_{i=1}^{n} \rho_{P_y b_t} + \sum_{i=1}^{n} (-1)^{i+1} |\Delta_{y,x_i}|_{\xi, \tau_i} \rho_{x_i y} \right],
\]
or
\[ \text{MSE} \left( \gamma_{y(x_i, \tau_i)} \right) = \frac{\theta \bar{Y}^2 C_y^2}{\Delta_{(x_i, \tau_i)}} \left[ \Delta_{(y, x_i, \tau_i)} \right]. \]  \hspace{1cm} (4.1.7)

**Corollary 5.1.1**

Let \( \gamma_{y(x_i, \tau_i)} \) be a mixture regression estimator based on one auxiliary variable and one auxiliary attribute given by,
\[
\gamma_{y(x_i, \tau_i)} = \bar{y} + \alpha_1^* (P_i - p_i) + \beta_1^* \left( \bar{x}_1 - \bar{x}_i \right)
\]
The mean square error of \( \gamma_{y(x_i, \tau_i)} \) is
\[
\text{MSE} \left( \gamma_{y(x_i, \tau_i)} \right) = \frac{\theta \bar{Y}^2 C_y^2}{\Delta_{(x_i, \tau_i)}} \left[ \Delta_{(y, x_i, \tau_i)} \right]. \] \hspace{1cm} (4.2.1)

**Corollary 5.1.2**

Let \( \gamma_{y(x_i, x_2, \tau_i)} \) be a mixture regression estimator based on two auxiliary variables and one auxiliary attribute given by,
\[
\gamma_{y(x_i, x_2, \tau_i)} = \bar{y} + \alpha_1^* (P_i - p_i) + \beta_1^* \left( \bar{x}_1 - \bar{x}_i \right) + \beta_2^* \left( \bar{x}_2 - \bar{x}_i \right)
\]
The mean square error of \( \gamma_{y(x_i, x_2, \tau_i)} \) is,
\[
\text{MSE} \left( \gamma_{y(x_i, x_2, \tau_i)} \right) = \frac{\theta \bar{Y}^2 C_y^2}{\Delta_{(x_i, x_2, \tau_i)}} \left[ \Delta_{(y, x_i, x_2, \tau_i)} \right]. \] \hspace{1cm} (4.2.2)

**Theorem 5.2**

Let \( \gamma_{y(x_1)} \) be a mixture regression estimator based on ‘m’ auxiliary attributes and single auxiliary variable given by:
\[
\gamma_{y(x_1)} = \bar{y} + \beta_1^* \left( \bar{x}_1 - \bar{x}_i \right) + \sum_{i=1}^{m} \alpha_i^* (P_i - p_i)
\] \hspace{1cm} (5.1.1)

The mean square error of \( \gamma_{y(x_1)} \) is,
\[
\text{MSE} \left( \gamma_{y(x_1)} \right) = \frac{\theta \bar{Y}^2 C_y^2}{\Delta_{(x_1)}} \left[ \Delta_{(y, x_1)} \right]. \] \hspace{1cm} (5.1.2)

**Proof:**

We may write
\[
\gamma_{y(x_1)} = \bar{y} + \beta_1^* \left( \bar{x}_1 - \bar{x}_i \right) + \sum_{i=1}^{m} \alpha_i^* (P_i - p_i)
\]
using the given results as,
\[ \gamma_{y(\bar{z})} - \bar{Y} = \bar{e}_y - \beta_1 \bar{e}_{x_1} - \sum_{i=1}^{m} \alpha_i \bar{e}_{\tau_i} \]

\[ \text{MSE} \left( \gamma_{y(\bar{z})} \right) = E \left[ \bar{e}_y - \sum_{i=1}^{m} \alpha_i \bar{e}_{\tau_i} - \beta_1 \bar{e}_{x_1} \right]^2, \quad (5.1.3) \]

Partially differentiating (5.1.3) w.r.t. \( \alpha_i \)'s and \( \beta_1 \), \( (i = 1, 2, \ldots, m) \) and equating the results to zero we obtain the optimum values of \( \alpha_i \)'s and \( \beta_1 \) by solving the \( \text{m+1} = q \) Normal equations given by,

\[ \Delta_{(\bar{z},x_i)} b^*_{q\times1} = \bar{Y}C_y d_{q\times1} \]

or

\[ b^* = \bar{Y}C_y \Delta_{(\bar{z},x_i)}^{-1} d, \quad \text{provided} \quad |\Delta| \neq 0 \quad (5.1.4) \]

or

\[ b^* = \bar{Y}C_y \frac{\text{Adj} \left( \Delta_{(\bar{z},x_i)} \right) d}{|\Delta_{(\bar{z},x_i)}|}. \quad (5.1.5) \]

or

\[
\begin{bmatrix}
P_1 C_1 \alpha_1 \\
P_2 C_2 \alpha_2 \\
\vdots \\
P_m C_m \alpha_m \\
\bar{X}C_{x_1} \beta_1
\end{bmatrix}
= \begin{bmatrix}
(-1)^{1+1} \left| \Delta_{y,\tau_1} \right|_{\bar{z},x_i} \\
(-1)^{2+1} \left| \Delta_{y,\tau_2} \right|_{\bar{z},x_i} \\
\vdots \\
(-1)^{m+1} \left| \Delta_{y,\tau_m} \right|_{\bar{z},x_i} \\
(-1)^{q+1} \left| \Delta_{y,x_1} \right|_{\bar{z},x_i}
\end{bmatrix}
\]

where \( q = m+1 \).

Rewriting (5.1.3) we get,

\[ \text{MSE} \left( \gamma_{y(\bar{z})} \right) = E \left[ \bar{e}_y - \sum_{i=1}^{m} \alpha_i \bar{e}_{\tau_i} - \beta_1 \bar{e}_{x_1} \right] \]

(5.1.7)

Substituting the values of \( \alpha_i \)'s and \( \beta \) in (5.1.7) and simplifying we get,

\[ \text{MSE} \left( \gamma_{y(\bar{z})} \right) = \theta \bar{Y}^2 C_y^2 \left[ 1 + \sum_{i=1}^{m} \frac{(-1)^{i+1}}{\left| \Delta_{(\bar{z},x_i)} \right|} P_{C_j} P_{\rho_{P_i}} \left| \Delta_{y,\tau_i} \right|_{\bar{z},x_i} \right] \]

\[ + \frac{(-1)^{q+1}}{\left| \Delta_{(\bar{z},x_i)} \right|} \bar{X}_{C_{x_1}} \rho_{x_1} \]
or

$$MSE\left( \gamma_{y(\tau_1, \tau_2, x_1)} \right) = \frac{\theta \bar{Y}^2 C_y^2}{|\Delta_{(\tau_1, \tau_2, x_1)}|} \left[ |\Delta_{(y, \tau_1, \tau_2, x_1)}| \right].$$

**Corollary 5.2.1**

Let $\gamma_{y(\tau_1, \tau_2, x_1)}$ be a mixture regression estimator based on two auxiliary attributes and single auxiliary variable given by:

$$\gamma_{y(\tau_1, \tau_2, x_1)} = \bar{Y} + \sum_{i=1}^{2} \alpha_i^* (P_i - p_i) + \beta_1^* (\bar{X}_1 - \bar{x}_1)$$

(5.2.1)

The mean square error of $\gamma_{y(\tau_1, \tau_2, x_1)}$ is,

$$MSE\left( \gamma_{y(\tau_1, \tau_2, x_1)} \right) = \frac{\theta \bar{Y}^2 C_y^2}{|\Delta_{(\tau_1, \tau_2, x_1)}|} \left[ |\Delta_{(y, \tau_1, \tau_2, x_1)}| \right].$$

(5.2.2)

**Note:**

We observe that the mean square errors of mixture ratio and mixture regression estimators discussed above are similar to each other.

### 6. APPLICATION TO REAL LIFE DATA

A study was conducted on a population of 237 patients taken from three teaching hospitals of Lahore and the following variables and attributes were considered:

- $Y = Weight$, $X_1 = Age$, $X_2 = Height$, $\tau_1 = 1 (male)$,
- $\tau_1 = 2 (female)$, $\tau_2 = 0 (illiterate)$, $\tau_2 = 1 (literate)$.

We are interested to estimate the population mean weight. The estimators proposed in section 4 and section 5 along with their mean square errors and comparison of relative efficiency follows:

| $N$ | $n$ | $\theta$ | $\bar{Y}$ | $C_y$ | $|\Delta_{(\tau_1, x_1)}|$ | $|\Delta_{(y, \tau_1, x_1)}|$ | $|\Delta_{(\tau_1, \tau_2, x_1)}|$ | $|\Delta_{(y, \tau_1, \tau_2, x_1)}|$ |
|-----|-----|---------|-------|------|----------------|----------------|----------------|----------------|
| 237 | 80  | .01672  | 74.557| 12.545| 0.7159         | 0.9505         | 0.9923         | 0.6856         |

**Correlations**

<table>
<thead>
<tr>
<th></th>
<th>$Y$</th>
<th>$X_1$</th>
<th>$X_2$</th>
<th>$\tau_1$</th>
<th>$\tau_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$Y$</td>
<td>.250</td>
<td>.493</td>
<td>.485</td>
<td>-.023</td>
<td></td>
</tr>
<tr>
<td>$X_1$</td>
<td></td>
<td>.096</td>
<td>.088</td>
<td>.138</td>
<td></td>
</tr>
<tr>
<td>$X_2$</td>
<td></td>
<td></td>
<td>.658</td>
<td>-.102</td>
<td></td>
</tr>
<tr>
<td>$\tau_1$</td>
<td></td>
<td></td>
<td></td>
<td>-.139</td>
<td></td>
</tr>
<tr>
<td>$\tau_2$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Relative Efficiency Comparison

<table>
<thead>
<tr>
<th>Estimator</th>
<th>$\phi^*_{(y_1, x_1)}$</th>
<th>$\phi^*_{(y_1, x_2)}$</th>
<th>$\phi^*_{(y_2, x_1)}$</th>
<th>$\phi^*_{(y_2, x_2)}$</th>
<th>$\phi^*_{(y_1, x_1, x_2)}$</th>
<th>$\phi^*_{(y_1, x_1, x_2)}$</th>
<th>$\phi^*_{(y_2, x_2, x_2)}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\phi^*_{(y_1, x_1)}$</td>
<td>100</td>
<td>986.1505</td>
<td>1270.71</td>
<td>1018.17</td>
<td>99.98</td>
<td>93.56</td>
<td>99.18</td>
</tr>
<tr>
<td>$\phi^*_{(y_1, x_2)}$</td>
<td>100</td>
<td>128.86</td>
<td>103.25</td>
<td>10.14</td>
<td>9.49</td>
<td>10.06</td>
<td></td>
</tr>
<tr>
<td>$\phi^*_{(y_2, x_1)}$</td>
<td>100</td>
<td>80.13</td>
<td>7.87</td>
<td>7.36</td>
<td>7.81</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\phi^*_{(y_2, x_2)}$</td>
<td>100</td>
<td>9.82</td>
<td>9.19</td>
<td>9.74</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\phi^*_{(y_1, x_1, x_2)}$</td>
<td>100</td>
<td>93.58</td>
<td>99.20</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\phi^*_{(y_1, x_1, x_2)}$</td>
<td>93.58</td>
<td>100</td>
<td>106.06</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\phi^*_{(y_2, x_2, x_2)}$</td>
<td>99.20</td>
<td>106.06</td>
<td>100</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
CONCLUSION

\( \phi^*_{(y,x_1,x_2)} \) is more efficient than \( \phi^*_{(y,x_1,x_2)}, \phi^*_{(y,x_1,x_2)} \) and \( \phi^*_{(y,x_1,x_2)} \).

\( \phi^*_{(y,x_1,x_2)} \) is more efficient than \( \phi^*_{(y,x_1,x_2)} \) and \( \phi^*_{(y,x_1,x_2)} \).

\( \phi^*_{(y,x_1,x_2,x_2)} \) is more efficient than \( \phi^*_{(y,x_1,x_2)}, \phi^*_{(y,x_1,x_2)}, \phi^*_{(y,x_2,x_2)} \) and \( \phi^*_{(y,x_2,x_2)} \).

\( \phi^*_{(y,x_1,x_2,x_2)} \) is more efficient than \( \phi^*_{(y,x_1,x_2)}, \phi^*_{(y,x_1,x_2)}, \phi^*_{(y,x_2,x_2)} \) and \( \phi^*_{(y,x_2,x_2)} \).

\( \phi^*_{(y,x_1,x_2,x_2)} \) is the most efficient estimator among all the estimators.

REFERENCES

COMPARISON OF THREE DIFFERENT METHODS FOR OUTLIER DETECTION IN BIOEQUIVALENCE STUDIES

A. Rasheed 1 and T. Ahmad2

1 Department of Research, Dow University of Health Sciences, Karachi, Pakistan. Email: abdur.rasheed@duhs.edu.pk
2 Center for Bioequivalence Studies at ICCBS, University of Karachi, Karachi, Pakistan. Email: tasneem_ahmad2001@yahoo.com

ABSTRACT

The significance of bioequivalence (BE) studies is rising due to large scale production and utilization of generic products all over the world. On some occasions in regulatory submission and at times during in-house evaluation of different formulations and dosage forms the identification of outliers becomes obligatory. Correct identification of outlying data in BE studies is substantial for deciding the two products either bioequivalence or bioinequivalence. For the detection of outliers in BE studies with the crossover designs different methods have been suggested in the literature. In the present work, we compared three tests; (i) Likelihood distance (LD) test (ii) estimated distance (ED) test and principal component analysis (PCA) test. In order to compare these tests, we used two-way and three-way crossover data sets available at US-FDA website. All tests were employed on linear and logarithmic scales. During the course of work it was found interesting and note-worthy that ED and principal component analysis tests are more sensitive than LD test and sensitivity of ED test persists even for log-transformed data.

KEYWORDS

Bioequivalence; outliers; Likelihood distance; estimated distance; principal component.

1. INTRODUCTION

The term that considered as the rate and extension in which an active molecule is absorbed and becomes available at the drug action site is known as bioavailability. When comparison of pharmacokinetic parameters related to bioavailability is made between two formulations, the phenomena referred as bioequivalence. One of the two drugs is considered as the reference. The major pharmacokinetic parameters employed for the bioavailability assessment are: Area under the curve (AUC), Peak concentration (C_{max}), Time for achieving peak concentration (T_{max}). Bioequivalent drugs are pharmaceutical equivalents that, when administrated in the similar molar dose, in the identical conditions, does not reveal significant statistical differences concerning bioavailability.

The most broadly used measure of bioequivalence is average bioequivalence. Average bioequivalence depends on the comparison of difference between formulations in the sense of mean and including the fact that the distributions of selected pharmacokinetic parameters may differ between two formulations in other distributional
Comparison of three different methods for outlier detection

characteristics [synopsis6]. According to Food and Drug Administration FDA (2001) guidelines, two drugs or formulations are declared to be bioequivalent if the 90% confidence interval of the ratio of geometric mean of pharmacokinetic parameters such as, AUC and $C_{\text{max}}$, lies within the pre-specified range (80%, 125%).

Sometime in regulatory submission, for evaluation of different formulations, the detection of outliers becomes mandatory. Correct identification and treatment of outlying data in BE studies is substantial for deciding the two products either bioequivalence or bioinequivalent.

During the analysis of bioequivalence the logarithmic transformation of pharmacokinetic parameters is recommended by the FDA (2001) under the fact that logarithmic transformation makes the distribution much closer to the normal.

Lund (1975) has developed a method for outlier detection in linear model, use of this method has also been suggested by the FDA but chow and Liu (2009) pointed out and proved that this method no longer appropriate for the crossover design due to correlated pharmacokinetic responses from the same subject.

For detection of outliers in bioequivalence studies, two procedures based on Cook’s likelihood distance and the estimates distance, were proposed by chow and Tse (1990). Enachescu and Enachescu (2009) used principal components to introduce a test for outlier detection in bioequivalence studies with crossover design.

Ramsay and Elkum (2005) compared different outlier detection methods proposed by chow and Tse (1990) and Liu and weng (1991) and with the help of the simulation study, he shown that estimated distance test performed better than other tests.

In this present work we compared three outlier detection tests, likelihood distance test, estimated distance test and principal component analysis test. We performed these three tests on 2X2 and 3X3 crossover data sets and observed the numbers of subjects were detected as outlie
ers in these bioequivalence data sets. This work is carried on linear scale and as well as logarithmic scale as recommended by the FDA (2001). We also performed a simulation study and observed the performance of each test.

2. THREE OUTLIER DETECTION TEST FOR CROSSOVER DESIGN

In a crossover design comparing f formulations of a drug, the model is given as

$$y_{ijl} = \mu + S_i + P_l + F_j + e_{ijl} \quad i = 1, ..., n \quad j, l = 1, ..., f$$

where $y_{ijl}$ is the response of the ith subject in lth period for jth formulation, $\mu$ is the overall mean, $P_l$ is the fixed effect of the lth period with $\sum P_l = 0$, $F_j$ is the fixed effect for the jth formulation with $\sum F_j = 0$, $S_i$ is the random effect for the ith subject, and $e_{ijl}$ is the random error in observing $y_{ijl}$. It is assumed that the $\{S_i\}$ and $\{e_{ijl}\}$ are independently and normally distributed with means zero and variances $\sigma_s^2$ and $\sigma_e^2$, respectively.
2.1 Principal Component Analysis

The objective of PCA is to discover or to reduce the dimensionality of the data set and identify new meaningful underlying variables. Enachescu and Enachescu (2009) has mentioned that for normally distributed observation \( U_i \), are independent \( \chi^2_{1,j} \) variables, where \( \lambda_i \) is called Eigen value denotes the variance of the i-th principal component. He also considered \( \sum_{i=1}^{p} \lambda_i \chi^2_{1,j} \) the weighted sum of square distance to zero of the projected data into principal factorial plane, with \( E\left( \sum_{i=1}^{p} \lambda_i \chi^2_{1,j} \right) = \sum_{i=1}^{p} \lambda_i = p \) and \( \text{Var}\left( \sum_{i=1}^{p} \lambda_i \chi^2_{1,j} \right) = 2\sum_{i=1}^{p} \lambda_i^2 \). Now the Observations with a square distance greater than \( m \) (the rule of \( 2\sigma \)) may be considered as outliers where \( m = p + 2 \sqrt{2\sum_{i=1}^{p} \lambda_i^2} \).

2.2 Likelihood Distance (LD) Test

Chow and Tse (1990) introduced the likelihood distance test for identifying outlier in a bioequivalence study where the null hypothesis assumes that there are no period and formulation effects. Now the model becomes

\[
y_{ij} = \mu + S_i + e_{ij} \quad i=1, ..., n \quad j=1, ..., f
\]

The parameters of interest are \( \mu, \sigma_s^2, \) and \( \sigma_e^2 \). Let \( \theta=(\theta_1, \theta_2, \theta_3)' \), where \( \theta_1 = \mu \), \( \theta_2 = \sigma_e^2 \), and \( \theta_3 = \sigma_s^2 + f\sigma_e^2 \). The maximum likelihood function is given by

\[
L(\theta) = -\frac{Nf}{2} \log(2\pi) - \frac{N}{2} \log\left( \theta_2\theta_3^{f-1} \right) - \frac{1}{2\theta_3} \sum_{i=1}^{N} \sum_{j=1}^{f} \left( Y_{ij} - \theta_1 \right)^2 - \frac{f}{2} \left( \frac{1}{\theta_2} - \frac{1}{\theta_3} \right) \sum_{i=1}^{N} (\bar{Y}_i - \theta_1)^2
\]

The maximum likelihood estimator \( \hat{\theta} \) of \( \theta \) is given by

\[
\hat{\theta}_1 = \bar{Y} = \frac{1}{nf} \sum_{i} \sum_{j} Y_{ij}, \quad \hat{\theta}_2 = \frac{1}{n(f-1)} \sum_{i} \sum_{j} (Y_{ij} - \bar{Y}_i)^2
\]

\[
\hat{\theta}_3 = \frac{f}{n} \sum_{i} \sum_{j} (\bar{Y}_i - \bar{Y})^2
\]

The likelihood distance statistic for the ith subject is twice the difference between the model log likelihood evaluated by using the estimates from all the subjects and from the estimates obtained after deleting the ith subject.

\[
LD_i(\hat{\theta}) = 2 \left[ L(\hat{\theta}) - L_i(\hat{\theta}_{-i}) \right]
\]
where \( \hat{\theta}_{-i} \) is the maximum likelihood estimator of \( \theta \) obtained by deleting the ith subject from the data. Asymptotically, as \( n \to \infty \), \( LD_i(\hat{\theta}) \) is distributed as a chi-square statistic with three degrees of freedom. The ith subject is called an outlier if \( LD_i(\hat{\theta}) > \chi^2(3) \)

### 2.3 Estimates Distance (ED) Test

Estimated distance statistics depends on the difference in the parameter estimates arising from the deletion of the ith observation, rather than on the difference in the log-likelihood. The estimated distance statistic is

\[
ED_i(\hat{\theta}) = f^2 = \left( \hat{\theta} - \hat{\theta}_{-i} \right)^T \hat{\Sigma}^{-1} \left( \hat{\theta} - \hat{\theta}_{-i} \right)
\]  

(5)

where \( \hat{\Sigma}^{-1} \) is the maximum likelihood estimator of the variance matrix

\[
\hat{\Sigma} = \begin{bmatrix}
\theta_3 / n & 0 & 0 \\
0 & 2\theta_2^2 / (n-1) & 0 \\
0 & 0 & 2\theta_3^2 \\
\end{bmatrix}
\]

Chow and Tse (1990) proved that, \( ED_i(\hat{\theta}) \) is asymptotically distributed as a chi-square variable with three degrees of freedom. Hence, the estimated distance test declares the ith subject as an outlier if \( ED_i(\hat{\theta}) > \chi^2(3) \).

### 3. APPLICATION AND RESULTS

#### 3.1 3x3 Crossover Design

In order to apply above defined test procedures for detecting outliers in BE studies, we found a BE data set on FDA website, where it is mentioned as data set 8. In that BE study there were 3 formulations (A, B and C), 3 periods and 3 sequences (ACB, BAC and CBA); it was a crossover design with equal number of sequence, periods and formulations. 1st Sequence contained 8 subjects, 2nd sequence 7 and 3rd had only 6 subjects. As only 21 subjects out of 24 subjects completed the study, now for the sake of our convenience we coded them as subject number from 1 to 21.

In figure 1, likelihood and estimated distances for LD and ED tests and observations squared distances for principal component test are presented for linear scale data. For LD and ED tests, subject is to be considered as outlier if \( LD_i(\hat{\theta}) > \chi^2(3) = 7.81473 \) and \( ED_i(\hat{\theta}) > \chi^2(3) = 7.81473 \) respectively and for PCA test, subject with squared distance greater than threshold value \( m = p + 2 \sqrt{\frac{3}{2} \sum_{i=1}^{3} \lambda_i^2} = 9.7934464 \) is considered as outlier.

Figure 2 depicts the results on logarithmic scale and for PCA test threshold value was 9.5963537.
Fig. I: For 3x3 crossover BE data, likelihood distance, estimated distance and observation squared distances for LD, ED and PCA tests respectively, on linear scale.

Fig. II: For 3x3 crossover BE data, likelihood distance, estimated distance and observation squared distances for LD, ED and PCA tests respectively, on logarithmic scale.

Table 1: Result of outlier detection from each method for 3x3 crossover data on linear and logarithmic scale

<table>
<thead>
<tr>
<th></th>
<th>BE data on linear scale</th>
<th>BE data on logarithmic scale</th>
</tr>
</thead>
<tbody>
<tr>
<td>LD test</td>
<td>ED test</td>
<td>PCA test</td>
</tr>
<tr>
<td>No outlier</td>
<td>Subj. 20 is outlier</td>
<td>Subj. 20 is outlier</td>
</tr>
</tbody>
</table>

From above figures and table it is evident that none of the subject was detected as an outlier from likelihood method, subject 20 was detected as an outlier from estimated distance method on both scales whereas principal component method detected it only on linear scale.
3.2 2x2 Crossover Design

We considered a BE data set from Chow and Liu (2009) this study was conducted with 24 healthy volunteers. The design of the BE study was 2x2 crossover i.e., there were two sequences (RT and TR) and two periods. Each of the two sequences contained 12 subjects. Five -50-mg tablets as test formulations or 5-mL of an oral suspension as a reference formulation were given to each of the subjects during each doing period.

In figure 3, likelihood, estimated distances and observations squared distances for likelihood distance test and estimated distance test and principal component methods respectively, are presented for linear scale data. For likelihood distance test, estimated distance test, the criterion for declaring a subject as an outlier is same as defined above and for PCA test, subject with squared distance greater than threshold value

\[
m = p + 2 \sqrt{2 \sum_{i=1}^{3} \lambda_i^2}
\]

6.7324548 is considered as outlier. Figure 2 shows the results on logarithmic scale. For PCA test threshold value was 6.5054357.

**Fig. III:** For 2X2 crossover BE data, likelihood distance, estimated distance and observation squared distances for LD, ED and PCA tests respectively, on linear scale.

**Fig. IV:** For 2X2 crossover BE data, likelihood distance, estimated distance and observation squared distances for LD, ED and PCA tests respectively, on logarithmic scale.
Table 2: Result of outlier detection from each method for 2x2 crossover data on linear and logarithmic scale

<table>
<thead>
<tr>
<th>BE data on linear scale</th>
<th>BE data on logarithmic scale</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>LD test</strong></td>
<td><strong>ED test</strong></td>
</tr>
<tr>
<td>No outlier</td>
<td>Subj. 13 and 19 are outliers</td>
</tr>
</tbody>
</table>

From above figures (III and IV) and table 2 it is evident that none of the subject was detected as an outlier from likelihood method, subjects 13 and 19 were detected as outliers from estimated distance method on linear scales and subjects 13 and 23 were detected as outlier on logarithmic scale, moreover, on logarithmic, subject 19 was not as outlier but very close to the threshold value. Whereas principal component analysis method detected subject 19 as an outlier on linear scale and on logarithmic scale subject 13 was an outlier, subjects 19 and 23 was very close to the threshold value but was not an outlier.

4. SIMULATION STUDY

For each combination of sample size=18, 24 and intrasubject variabilities = 20, 40, a total of 100 data sets of AUC values were generated from the statistical model for 2X2 crossover design under the normality assumption. For the sake of our convenience we there were no period and carryover effects. The true test and reference means were both chosen to be 100. The results of simulations study are given in the table 3 and 4 on linear and logarithmic scales respectively. For example on linear scale, outliers for sample size 24 and intrasubject variabilities 20, the estimated distance test performed better than likelihood distance test in 97% and better than PCA test in 82% under the same condition estimated distance test performed at least as well as likelihood distance test in 100% and at least as well as PCA test in 100%.

In Table 4, for sample size 24 and intrasubject variabilities 20, the estimated distance test performed better than likelihood distance test in 90% and better than PCA test in 62% under the same condition estimated distance test performed at least as well as likelihood distance test in 100% and at least as well as PCA test in 99%.

Table 3:

The simulations percentage where estimated distance test performed better than other tests on linear scale, in terms of detecting outliers more frequently, and the percentage in which it was at least as good.

<table>
<thead>
<tr>
<th>Linear scale</th>
<th>Sample size</th>
<th>Intrasubject variabilities</th>
<th>ED&gt;LD (ED≥LD)</th>
<th>ED&gt;PCA (ED≥PCA)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>24</td>
<td>20</td>
<td>97(100)</td>
<td>82(100)</td>
</tr>
<tr>
<td></td>
<td>18</td>
<td></td>
<td>94(100)</td>
<td>71(100)</td>
</tr>
<tr>
<td></td>
<td>24</td>
<td>40</td>
<td>95(100)</td>
<td>77(100)</td>
</tr>
<tr>
<td></td>
<td>18</td>
<td></td>
<td>97(100)</td>
<td>78(100)</td>
</tr>
</tbody>
</table>
Table 4:
The simulations percentage where estimated distance test performed better than other tests on logarithmic scale, in terms of detecting outliers more frequently, and the percentage in which it was at least as good.

<table>
<thead>
<tr>
<th>Sample size</th>
<th>Log scale</th>
<th>Intrasubject variabilities</th>
<th>ED&gt;LD (ED&gt;LD)</th>
<th>ED&gt;PCA (ED&gt;PCA)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>24</td>
<td>20</td>
<td>90(100)</td>
<td>62(99)</td>
<td></td>
</tr>
<tr>
<td>18</td>
<td></td>
<td>86(100)</td>
<td>61(98)</td>
<td></td>
</tr>
<tr>
<td>24</td>
<td>40</td>
<td>80(100)</td>
<td>61(93)</td>
<td></td>
</tr>
<tr>
<td>18</td>
<td></td>
<td>84(100)</td>
<td>63(99)</td>
<td></td>
</tr>
</tbody>
</table>

5. CONCLUSION

Findings of our work powerfully propose that estimated distance method is sensitive and superior to likelihood distance and principal component analysis methods. Sensitivity of estimated distance persists even for log-transformed data and for both 2x2 and 3x3 crossover bioequivalence data sets. It can also obvious from above findings that principal component analysis method is much superior to likelihood distance test as it did not detect any subject as an outlier for both BE data sets on any of scale. And also the results of our simulation powerfully advocate that the Estimates distance test is superior to likelihood distance test and PCA method for identifying outliers in bioequivalence studies. By using simulation study Ramsay and Elkum (2005) has also shown that estimated distance test more sensitive than other outlier detection tests, but no bioequivalence study data set was employed in that research.

REFERENCES

ON DISCORDANCE TEST FOR THE WRAPPED NORMAL DATA

Adzhar Rambli\textsuperscript{1}, Ibrahim Mohamed\textsuperscript{1}, Abdul Ghapor Hussin\textsuperscript{2} and Safwati Ibrahim\textsuperscript{1}

\textsuperscript{1} Institute of Mathematical Sciences, University of Malaya, Kuala Lumpur, Malaysia. Email: adzfranc@yahoo.com, imohamed@um.edu.my, safwati@um.edu.my
\textsuperscript{2} Centre for Foundation studies in Sciences, University of Malaya, Kuala Lumpur, Malaysia. Email: ghapor@um.edu.my

ABSTRACT

This paper focuses on detecting outliers in the circular data which follow the wrapped normal distribution. We consider four discordance tests based on M, C, D and A Statistics. The cut-off points of the four tests are obtained and the performance of the detection procedures is studied via simulations. In general, we show that the discordance test based on the A statistic outperforms the other tests in all cases. For illustration, the Kuantan wind direction data set is considered.

KEYWORDS

Discordance; Circular; wrapped normal distribution; outlier; A statistics.

INTRODUCTION

Circular data can be visualized as being distributed on the circumference of a unit circle in the range of 0 to 2\pi radian. The data are commonly found in many scientific fields such as meteorology and biology where researchers are interested in studying direction of wind and direction of movement of animals, respectively. Due to the bounded range property of circular variables, special methods such as circular descriptive statistics, circular plots and goodness of fit tests are required to describe and model such data.

Several authors have comprehensively discussed the circular distributions including Jammalamadaka and SenGupta (2001), Mardia (1972) and Fisher (1993). Various distributions are available for circular data, for example, uniform distribution, wrapped Cauchy distribution, wrapped normal distribution, cardioid distribution, and others. Jammalamadaka and SenGupta (2001) reviewed the wrapped \( \alpha \) stable distribution with the wrapped Cauchy and the wrapped normal distributions as the special cases. On the other hand, several bivariate circular distributions exist, such as the bivariate von Mises distribution, wrapped bivariate normal distribution and circular-linear distribution. The von Mises distribution (also known as the circular normal distribution) is the most commonly used and is a continuous probability distribution on a circle. The von Mises distribution may be thought of as a close approximation to the wrapped normal distribution, which is the circular analogue of the normal distribution.
As in the linear case, the existence of outliers in circular data is expected to affect the estimation of parameters and weaken the accuracy of forecast. Thus, it is very important that methods of identifying outliers in circular data are developed for proper handling of the data. Graphical and numerical methods are the most common tools used in investigating the existence of outliers in circular data. We consider four discordance tests to detect possible outlier in the circular data based on the $M$, $C$, $D$ and $A$ statistics. The later is proposed by Abuzaid (2009) which has been shown to perform better than other methods for data from the $VM$ distribution, except for small sample size. In this paper, we apply the tests on data coming from wrapped normal ($WN$) distributions.

With that view in mind, this paper is organized as follows: Section 2 describes properties of the wrapped normal distribution. Section 3 reviews several discordance tests to detect the existence of outlier in circular univariate data. In Section 4, we obtain the cut off-points and study the performance of each statistics by simulation studies to detect outlier in circular data that come from the wrapped normal distribution. We then apply the statistics on the real data set obtained from the Malaysian Meteorological Service Department in the last section.

WRAPPED NORMAL DISTRIBUTION

Jammalamadaka and SenGupta (2001) discussed the general wrapped $\alpha$-stable distribution which is constructed by using the characteristic function of the $\alpha$-stable of a real line. The characteristic function as given by Lukasc (1970) is

$$
\psi(t) = \begin{cases} 
\exp \left\{ -\tau^\alpha |t|^\alpha \left[ 1 - i \beta \text{sgn}(t) \tan \left( \frac{\alpha \pi}{2} \right) \right] + i \mu \right\}, & \text{if } \alpha \in (0,1) \cup (1,2], \\
\exp(-|t|) + i \mu, & \text{if } \alpha = 1,
\end{cases}
$$

where $\tau \geq 0$, $|\beta| \leq 1$, $0 < \alpha \leq 2$, while $\mu$ is a real number. The density function of a wrapped $\alpha$-stable random variable for $\theta \in [0,2\pi)$ is given by

$$
f(\theta) = \frac{1}{2\pi} + \frac{1}{\pi} \sum_{k=1}^{\alpha} \exp \left\{ -\tau^\alpha k^\alpha \right\} \cos \left\{ k(\theta - \mu) - \tau^\alpha k^\alpha \beta \tan \left( \frac{\alpha \pi}{2} \right) \right\},
$$

when $\alpha \in (0,1) \cup (1,2]$, with $\mu$ conveniently redefined as $\mu \pmod{2\pi}$. Note that although there is generally no closed form expression for the density of an $\alpha$-stable distribution on the real line, we are able to write such density for the wrapped case, at least as an infinite series.

The particular case corresponding to $\beta = 0$ gives us the symmetric wrapped stable (SWS) family of circular densities, which we will simply refer to as wrapped stable (WS), given by

$$
f(\theta) = \frac{1}{2\pi} + \frac{1}{\pi} \sum_{k=1}^{\infty} \rho^k \cos \left\{ k(\theta - \mu) \right\},
$$
where $\rho = \exp\left(-\tau^\alpha\right)$. We shall denote such distributions as $WS(\alpha, \rho, \mu)$. The special case with $\alpha = 2$ and $\beta = 0$ gives us the wrapped normal density with $\rho = \exp\left(-\frac{\sigma^2}{2}\right)$. When $\alpha = 1$ and $\beta = 0$, it gives us the wrapped Cauchy density with $\rho = \exp\left(-\tau\right)$.

A wrapped normal distribution is obtained by wrapping a normal distribution around a unit circle. The normal distribution is denoted by $N(\mu_L, \sigma_L^2)$ where $\mu_L$ is the mean and $\sigma_L^2$ is the variance while the WN distribution is denoted by $WN(\mu, \rho)$, where $\mu$ is the mean direction and $\rho$ is the measure of concentration parameter. Its probability distribution function is given by

$$f(\theta) = \frac{1}{\sigma\sqrt{2\pi}} \sum_{k=-\infty}^{\infty} \exp\left[\frac{-(\theta - \mu - 2k\pi)^2}{2\sigma^2}\right]$$ (1)

where $\sigma^2$ is the circular variance. From Whittaker and Watson (1944), an alternate and more useful representation of this density is

$$f(\theta) = \frac{1}{2\pi} \left(1 + 2 \sum_{k=1}^{\infty} \rho^k \cos k(\theta - \mu)\right) \quad 0 \leq 0 < 2\pi, \quad 0 \leq \rho \leq 1.$$

The distribution is unimodal and symmetric about the value $\theta = \mu$. Unlike the von Mises distribution, the WN distribution possesses the additive property, that is, the convolution of two WN variables is also WN. Specifically, if $\theta_1 \sim WN(\mu_1, \rho_1)$ and $\theta_2 \sim WN(\mu_2, \rho_2)$ are independent, then $\theta_1 + \theta_2 \sim WN(\mu_1 + \mu_2, \rho_1 \rho_2)$ (see Jammalamadaka and SenGupta (2001)). We study the case when the sample data follows the wrapped normal distribution with the mean direction $\mu$ and the concentration parameter $\kappa$.

In deciding whether a circular data set follows the von Mises ($VM$) distribution or the wrapped normal ($WN$) distribution, Kent (1976) highlighted the fact that both distributions are hardly distinguishable for $\kappa < 0.1$ or $\kappa > 10$. In this case, Kendall (1974) noted that for any analytical, computational and statistical purposes, the WN distribution is more convenient to be used in some cases and the $VM$ distribution in other cases. Collet and Lewis (1981) concluded that a minimum sample size required in order to distinguish the two distributions is 200 via the classical discriminant approach.

**DISCORDANCE TESTS IN DETECTING OUTLIERS**

Suppose $\theta_1, \theta_2, \ldots, \theta_n$ are (i.i.d) circular observations located on the circumference of a unit circle. We consider four discordance tests based on the $C$, $D$, $M$, and $A$ statistics to identify outliers in a univariate circular sample from the $WN$ distribution.
i) C Statistic

The mean resultant length of circular data set is given by \( R = \frac{R}{n} \), where \( R = \sqrt{C^2 + S^2} \) such that \( C = \sum_{i=1}^{n} \cos \theta_i \) and \( S = \sum_{i=1}^{n} \sin \theta_i \). By omitting the \( i \)th observation, the mean resultant length is given by \( R_{(-i)} = \frac{R_{(-i)}}{n-1} \). Therefore,

\[
C = \max_i \left\{ \frac{R_{(-1)} - R}{R} \right\},
\]

(2)
can be considered as a test statistic. Values of \( C \) statistic will then be compared with the cut-off points for the corresponding sample size \( n \) and estimated concentration parameter \( \kappa \). If \( C \) is larger than the cut-off point, we reject the null hypothesis so that the \( i \)th observation is identified as an outlier.

ii) D Statistic

The \( D \) statistic uses the relative arc length based on the ordered observation of a circular sample \( \theta_{(1)}, \theta_{(2)}, \ldots, \theta_{(n)} \). Let \( T_i \) be the arc length between consecutive observations given by \( T_i = \theta_{(i+1)} - \theta_{(i)} \), \( i = 1, 2, \ldots, n \) and \( T_n = 2\pi - \theta_{(n)} + \theta_{(1)} \). Define \( D_i = \frac{T_i}{T_{i-1}} \), \( i = 1, 2, \ldots, n \) and \( T_0 \equiv T_n \). Let \( D_k = \frac{T_k}{T_{k-1}} \) corresponds to the greatest arc containing a single observation \( \theta_k \). Note that \( D_k \) is two tailed. Collett (1980) suggested working in terms of

\[
D = \min(D_k, D_k^{-1})
\]

(3)
where \( 0 < D < 1 \). The observation \( \theta_k \) can be considered as an outlier if the value of \( D \) is larger than the cut-off point.

iii) M Statistic

Mardia (1975) suggested a statistic of discordancy which is given by

\[
M' = \min_i \left\{ \frac{n-1-R_{(-i)}}{n-R} \right\}.
\]

Later, Collett (1980) reformulated the \( M' \) statistic in terms of

\[
M = 1 - M' = \max_i \left\{ \frac{R_{(-i)} - R + 1}{n-R} \right\} = \frac{R_y - R + 1}{n-R},
\]

(4)
where \( R_y = \max_i \left\{ R_{(-i)} \right\} \). He stated the asymptotic distribution of the \( M \) statistic for large values of \( \kappa \). As the value of \( \kappa \) increases, the von Mises distribution will be approximated by a standard normal distribution. On the other hand, the \( M \) statistic can
be approximated by \( \frac{n(b^*)^2}{n-1} \), where \( b^* \) is the test statistic used to identify discordancy in normal data. Percentage points for \( b^* \) are given in Pearson and Hartley (1966).

iv) A Statistic

Rao (1969) defined the circular distance between \( \theta_i \) and \( \theta_j \) as

\[
d_{ij} = 1 - \cos(\theta_i - \theta_j)
\]

where \( d_{ij} \) is a monotone increasing function of \((\theta_i - \theta_j)\) and \( d_{ij} \in [0, 2] \). The summation of all circular distances of the point of interest \( \theta_j \) to all other points is given by

\[
D_j = \sum_{i=1}^{n} (1 - \cos(\theta_i - \theta_j)), \; i = 1, 2, ..., n
\]

If the observation \( \theta_j \) is an outlier, then the value of \( D_j \) will increase. Thus, the average circular distance given by \( \frac{D_j}{n-1} \) can be used to identify possible outliers in the circular sample. Abuzaid (2009) proposed the A statistic as

\[
A = \max_{j} \left\{ \frac{D_j}{2(n-1)} \right\}, \; j = 1, 2, ..., n
\]  

(5)

where \( A \in [0, 1] \) is a linear measure. The average circular distance is divided by 2 in order to standardize the values of statistic \( A \). The proposed statistic is based on the relative decrease in the summation of circular distances by omitting the point of interest \( \theta_j \).

### CUT-OFF POINTS OF THE DISCORDANCE TESTS

The interest is on the use of the discordance tests to detect outlier in data generated from the wrapped normal distribution. Firstly, we have to obtain the cut-off points for each test. Thus, we design a simulation study to find the percentage points of the null distribution of no outliers in the circular data set. We consider eleven values of measure of concentration parameter in the range of 0.1 to 0.975 and different sample sizes from 5 to 150. For each combination of \( n \) and \( \rho \), we generate sample from \( WN(\mu = 0, \rho) \). All the test statistics in each generated random sample are calculated using statistics (2)-(5) respectively. We wish to estimate the percentage points of the discordance tests at the 10\%, 5\% and 1\% upper percentiles when no outlier presents in the sample.

Tables 1-4 show the cut-off points of the four tests. Two main results are observed. Firstly, as the measure of concentration parameter increases, the cut-off points decreases for the three levels of percentiles. This is expected as the circular data are more concentrated with larger \( \rho \) resulting in a smaller difference between two largest values of
the statistics. Secondly, as the sample size increases, the value of the cut-off point decreases. Again, this should be true as the sample size increases, the distance between the circular observations in circular plot become smaller.

**PERFORMANCE OF THE DISCORDANCE TESTS**

Collett (1980) applied selected measures to test the performances of several statistics to detect an outlier in circular sample. Here, we use similar measures to compare the performance of the tests. David (1970, p. 185), and Barnett and Lewis (1978, pp. 64–68), stated that a good test should have: (i) a high power function; (ii) a high probability of identifying a contaminating value as an outlier when it is in fact an extreme value, where an extreme value is defined as a point with the maximum circular deviation; and (iii) a low probability of wrongly identifying a good observation as discordant.

Let \( P_1 = 1 - \beta \) be the power function where \( \beta \) is the Type-II error; \( P_3 \) the probability that the contaminant point is an extreme point and is identified as discordant; and \( P_5 \) the probability that the contaminant point is identified as discordant given that it is an extreme point. A good test is expected to have (i) high \( P_1 \), (ii) high \( P_5 \), and (iii) low \( P_1 - P_3 \).

To study the performance of all discordance tests, we use 2000 simulations for different sizes of \( n \) and \( \rho \). The samples are generated in such a way that \((n-1)\) of the observations come from the \( WN(\alpha,\rho) \) and one observation from \( WN(\alpha + \lambda \pi,\rho) \), where \( \lambda \) is the degree of contamination and \( 0 \leq \lambda \leq 1 \). The \( C, D, M \) and \( A \) statistics in each random sample are then calculated.

Figures 1-3 illustrate the power of performance of the tests for different cases. Three main results are observed. Firstly, in Figures 1(a) and (b), for \( n = 50 \) and \( \rho = 0.90 \), the test based on the \( A \) statistic performs better than the others for all contamination levels \( \lambda \) since the \( P_1 \) and \( P_5 \) curves are always greater than the others. As in Figure 1 (c), the four tests have low \( P_1-P3 \) since the curves are almost 0. Secondly, from Figure 1(a) and Figure 2(a), \( P_1 \) reaches the value 1 when \( \lambda = 0.6 \) and \( \lambda = 0.8 \) respectively. Similar results are observed for \( P_5 \) as shown in Figure 1(b) and Figure 2(b). These suggest that as \( \rho \) gets larger, the four tests show better performance of detecting outliers at lower contamination level. Thirdly, from Figure 1(a) and Figure 3(a), when \( n \) becomes smaller, the test based on the \( A \) statistic performs better than that of the \( C \) and \( D \) statistics, but performs much better than that of the \( M \) statistic. Similar results are observed for \( P_5 \) as shown in Figure 1(b) and Figure 3(b).

The four discordance methods have been investigated for the case when the data come from the von Mises distribution by Abuzaid et al. (2009). The cut-off points for tests based on the \( C \) and \( D \) statistics can be obtained from Collet (1980) while that of the \( M \) and \( A \) statistics are available in Mardia (1975) and Abuzaid (2009) respectively. In summary, the result for the \( VM \) and \( WN \) are similar except for the case when \( n \) is small. In this particular case, the test based on the \( A \) statistic performs better than others in terms of \( P_1 \) and \( P_5 \) for the \( WN \) distribution but the test based on the \( M \) statistic performs better for the \( VM \) distribution.
APPLICATION

We consider the Kuantan wind direction data measured in unit radian from the year 1999 to 2008 as tabulated in Table 5 obtained from the Malaysian Meteorological Services Department. Table 6 gives the values of circular descriptive statistics for the data. The mean direction $\mu$ is 84.65° and the concentration parameter for this data is 3.33. We can conclude that the data sets are concentrated in the east direction. The circular plot of the data is given in Figure 4. We notice that there is one observation located a bit separated from the rest. Here, we have $n = 10$ and $\rho = 0.88$. Table 7 gives the value of the test statistics, the cut-off point for $n = 10$ and $\rho = 0.9$, as well as the decision for each statistic. It can be seen that the four tests do not identify the outlying observation as an outlier. Note, however, that the values of test statistics are very close to their respective cut-off points. Thus, it warrants further investigation on the observation.

CONCLUSION

In this paper, we have reviewed four discordance tests to identify the existence of outlier in circular data. The cut-off points for tests based on the $C$, $M$, $D$ and $A$ statistics for the wrapped normal distribution are obtained via simulation studies. We have compared the performance of the tests for the VM and the WN distributions. In general, the test based on the $A$ statistic outperforms the other tests. As an illustration, we apply the statistics to identify the existence of outlier on the Kuantan wind direction data.

REFERENCES


### Table 1:

Table of cut-off points for the test based on the $C$ statistic

<table>
<thead>
<tr>
<th>$n$</th>
<th>Level of percentile</th>
<th>$\rho$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>0.1</td>
</tr>
<tr>
<td>5</td>
<td>10%</td>
<td>1.711</td>
</tr>
<tr>
<td></td>
<td>5%</td>
<td>2.379</td>
</tr>
<tr>
<td></td>
<td>1%</td>
<td>4.867</td>
</tr>
<tr>
<td>10</td>
<td>10%</td>
<td>1.114</td>
</tr>
<tr>
<td></td>
<td>5%</td>
<td>1.542</td>
</tr>
<tr>
<td></td>
<td>1%</td>
<td>3.668</td>
</tr>
<tr>
<td>30</td>
<td>10%</td>
<td>0.551</td>
</tr>
<tr>
<td></td>
<td>5%</td>
<td>0.739</td>
</tr>
<tr>
<td></td>
<td>1%</td>
<td>1.398</td>
</tr>
<tr>
<td>50</td>
<td>10%</td>
<td>0.365</td>
</tr>
<tr>
<td></td>
<td>5%</td>
<td>0.528</td>
</tr>
<tr>
<td></td>
<td>1%</td>
<td>1.398</td>
</tr>
<tr>
<td>70</td>
<td>10%</td>
<td>0.286</td>
</tr>
<tr>
<td></td>
<td>5%</td>
<td>0.409</td>
</tr>
<tr>
<td></td>
<td>1%</td>
<td>0.954</td>
</tr>
<tr>
<td>90</td>
<td>10%</td>
<td>0.222</td>
</tr>
<tr>
<td></td>
<td>5%</td>
<td>0.321</td>
</tr>
<tr>
<td></td>
<td>1%</td>
<td>0.694</td>
</tr>
<tr>
<td>100</td>
<td>10%</td>
<td>0.202</td>
</tr>
<tr>
<td></td>
<td>5%</td>
<td>0.283</td>
</tr>
<tr>
<td></td>
<td>1%</td>
<td>0.544</td>
</tr>
<tr>
<td>150</td>
<td>10%</td>
<td>0.130</td>
</tr>
<tr>
<td></td>
<td>5%</td>
<td>0.177</td>
</tr>
<tr>
<td></td>
<td>1%</td>
<td>0.378</td>
</tr>
</tbody>
</table>
Table 2:
Table of cut-off points for the test based on the *M* statistic

<table>
<thead>
<tr>
<th>n</th>
<th>Level of percentile</th>
<th>( \rho )</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>0.1</td>
</tr>
<tr>
<td>5</td>
<td>10%</td>
<td>0.762</td>
</tr>
<tr>
<td></td>
<td>5%</td>
<td>0.838</td>
</tr>
<tr>
<td></td>
<td>1%</td>
<td>0.934</td>
</tr>
<tr>
<td>10</td>
<td>10%</td>
<td>0.349</td>
</tr>
<tr>
<td></td>
<td>5%</td>
<td>0.389</td>
</tr>
<tr>
<td></td>
<td>1%</td>
<td>0.487</td>
</tr>
<tr>
<td>30</td>
<td>10%</td>
<td>0.096</td>
</tr>
<tr>
<td></td>
<td>5%</td>
<td>0.102</td>
</tr>
<tr>
<td></td>
<td>1%</td>
<td>0.116</td>
</tr>
<tr>
<td>50</td>
<td>10%</td>
<td>0.054</td>
</tr>
<tr>
<td></td>
<td>5%</td>
<td>0.057</td>
</tr>
<tr>
<td></td>
<td>1%</td>
<td>0.061</td>
</tr>
<tr>
<td>70</td>
<td>10%</td>
<td>0.037</td>
</tr>
<tr>
<td></td>
<td>5%</td>
<td>0.038</td>
</tr>
<tr>
<td></td>
<td>1%</td>
<td>0.041</td>
</tr>
<tr>
<td>90</td>
<td>10%</td>
<td>0.028</td>
</tr>
<tr>
<td></td>
<td>5%</td>
<td>0.029</td>
</tr>
<tr>
<td></td>
<td>1%</td>
<td>0.031</td>
</tr>
<tr>
<td>100</td>
<td>10%</td>
<td>0.025</td>
</tr>
<tr>
<td></td>
<td>5%</td>
<td>0.026</td>
</tr>
<tr>
<td></td>
<td>1%</td>
<td>0.028</td>
</tr>
<tr>
<td>150</td>
<td>10%</td>
<td>0.016</td>
</tr>
<tr>
<td></td>
<td>5%</td>
<td>0.017</td>
</tr>
<tr>
<td></td>
<td>1%</td>
<td>0.018</td>
</tr>
</tbody>
</table>
Table 3:
Table of cut-off points for the test based on the the $D$ statistic

<table>
<thead>
<tr>
<th>n</th>
<th>Level of percentile</th>
<th>0.1</th>
<th>0.2</th>
<th>0.4</th>
<th>0.6</th>
<th>0.8</th>
<th>0.9</th>
<th>0.95</th>
<th>0.975</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>10%</td>
<td>0.872</td>
<td>0.857</td>
<td>0.811</td>
<td>0.587</td>
<td>0.277</td>
<td>0.171</td>
<td>0.105</td>
<td>0.071</td>
</tr>
<tr>
<td></td>
<td>5%</td>
<td>0.937</td>
<td>0.931</td>
<td>0.898</td>
<td>0.737</td>
<td>0.337</td>
<td>0.198</td>
<td>0.127</td>
<td>0.082</td>
</tr>
<tr>
<td></td>
<td>1%</td>
<td>0.990</td>
<td>0.984</td>
<td>0.984</td>
<td>0.909</td>
<td>0.487</td>
<td>0.275</td>
<td>0.169</td>
<td>0.109</td>
</tr>
<tr>
<td>10</td>
<td>10%</td>
<td>0.857</td>
<td>0.872</td>
<td>0.814</td>
<td>0.611</td>
<td>0.268</td>
<td>0.150</td>
<td>0.094</td>
<td>0.060</td>
</tr>
<tr>
<td></td>
<td>5%</td>
<td>0.924</td>
<td>0.933</td>
<td>0.911</td>
<td>0.762</td>
<td>0.323</td>
<td>0.185</td>
<td>0.114</td>
<td>0.072</td>
</tr>
<tr>
<td></td>
<td>1%</td>
<td>0.980</td>
<td>0.986</td>
<td>0.980</td>
<td>0.930</td>
<td>0.504</td>
<td>0.246</td>
<td>0.150</td>
<td>0.099</td>
</tr>
<tr>
<td>30</td>
<td>10%</td>
<td>0.872</td>
<td>0.875</td>
<td>0.814</td>
<td>0.695</td>
<td>0.259</td>
<td>0.132</td>
<td>0.078</td>
<td>0.051</td>
</tr>
<tr>
<td></td>
<td>5%</td>
<td>0.937</td>
<td>0.926</td>
<td>0.902</td>
<td>0.847</td>
<td>0.335</td>
<td>0.165</td>
<td>0.097</td>
<td>0.061</td>
</tr>
<tr>
<td></td>
<td>1%</td>
<td>0.988</td>
<td>0.985</td>
<td>0.981</td>
<td>0.966</td>
<td>0.494</td>
<td>0.226</td>
<td>0.133</td>
<td>0.093</td>
</tr>
<tr>
<td>50</td>
<td>10%</td>
<td>0.873</td>
<td>0.864</td>
<td>0.849</td>
<td>0.750</td>
<td>0.279</td>
<td>0.129</td>
<td>0.076</td>
<td>0.048</td>
</tr>
<tr>
<td></td>
<td>5%</td>
<td>0.938</td>
<td>0.939</td>
<td>0.923</td>
<td>0.857</td>
<td>0.358</td>
<td>0.159</td>
<td>0.089</td>
<td>0.059</td>
</tr>
<tr>
<td></td>
<td>1%</td>
<td>0.981</td>
<td>0.988</td>
<td>0.986</td>
<td>0.965</td>
<td>0.582</td>
<td>0.214</td>
<td>0.122</td>
<td>0.081</td>
</tr>
<tr>
<td>70</td>
<td>10%</td>
<td>0.873</td>
<td>0.869</td>
<td>0.861</td>
<td>0.758</td>
<td>0.252</td>
<td>0.125</td>
<td>0.073</td>
<td>0.046</td>
</tr>
<tr>
<td></td>
<td>5%</td>
<td>0.931</td>
<td>0.939</td>
<td>0.934</td>
<td>0.876</td>
<td>0.334</td>
<td>0.152</td>
<td>0.089</td>
<td>0.055</td>
</tr>
<tr>
<td></td>
<td>1%</td>
<td>0.989</td>
<td>0.985</td>
<td>0.985</td>
<td>0.965</td>
<td>0.555</td>
<td>0.218</td>
<td>0.127</td>
<td>0.075</td>
</tr>
<tr>
<td>90</td>
<td>10%</td>
<td>0.857</td>
<td>0.857</td>
<td>0.845</td>
<td>0.791</td>
<td>0.259</td>
<td>0.130</td>
<td>0.072</td>
<td>0.046</td>
</tr>
<tr>
<td></td>
<td>5%</td>
<td>0.933</td>
<td>0.930</td>
<td>0.923</td>
<td>0.893</td>
<td>0.324</td>
<td>0.156</td>
<td>0.089</td>
<td>0.056</td>
</tr>
<tr>
<td></td>
<td>1%</td>
<td>0.989</td>
<td>0.979</td>
<td>0.991</td>
<td>0.968</td>
<td>0.517</td>
<td>0.225</td>
<td>0.124</td>
<td>0.077</td>
</tr>
<tr>
<td>100</td>
<td>10%</td>
<td>0.864</td>
<td>0.861</td>
<td>0.870</td>
<td>0.779</td>
<td>0.264</td>
<td>0.127</td>
<td>0.071</td>
<td>0.045</td>
</tr>
<tr>
<td></td>
<td>5%</td>
<td>0.930</td>
<td>0.930</td>
<td>0.930</td>
<td>0.882</td>
<td>0.341</td>
<td>0.160</td>
<td>0.084</td>
<td>0.054</td>
</tr>
<tr>
<td></td>
<td>1%</td>
<td>0.984</td>
<td>0.984</td>
<td>0.980</td>
<td>0.974</td>
<td>0.541</td>
<td>0.219</td>
<td>0.120</td>
<td>0.075</td>
</tr>
<tr>
<td>150</td>
<td>10%</td>
<td>0.861</td>
<td>0.869</td>
<td>0.863</td>
<td>0.819</td>
<td>0.269</td>
<td>0.118</td>
<td>0.070</td>
<td>0.044</td>
</tr>
<tr>
<td></td>
<td>5%</td>
<td>0.922</td>
<td>0.928</td>
<td>0.932</td>
<td>0.905</td>
<td>0.353</td>
<td>0.142</td>
<td>0.086</td>
<td>0.053</td>
</tr>
<tr>
<td></td>
<td>1%</td>
<td>0.981</td>
<td>0.986</td>
<td>0.985</td>
<td>0.982</td>
<td>0.550</td>
<td>0.200</td>
<td>0.123</td>
<td>0.070</td>
</tr>
</tbody>
</table>
Table 4: Table of cut-off points for the test based on the $A$ statistic

<table>
<thead>
<tr>
<th>$n$</th>
<th>Level of percentile</th>
<th>$\rho$</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>10%</td>
<td>0.1</td>
<td>0.2</td>
<td>0.4</td>
<td>0.6</td>
<td>0.8</td>
<td>0.9</td>
<td>0.95</td>
</tr>
<tr>
<td>5</td>
<td></td>
<td>0.924</td>
<td>0.928</td>
<td>0.920</td>
<td>0.892</td>
<td>0.726</td>
<td>0.549</td>
<td>0.395</td>
</tr>
<tr>
<td></td>
<td>5%</td>
<td>0.944</td>
<td>0.948</td>
<td>0.946</td>
<td>0.927</td>
<td>0.783</td>
<td>0.592</td>
<td>0.438</td>
</tr>
<tr>
<td></td>
<td>1%</td>
<td>0.971</td>
<td>0.978</td>
<td>0.979</td>
<td>0.962</td>
<td>0.879</td>
<td>0.705</td>
<td>0.509</td>
</tr>
<tr>
<td>10</td>
<td></td>
<td>0.859</td>
<td>0.872</td>
<td>0.893</td>
<td>0.888</td>
<td>0.751</td>
<td>0.569</td>
<td>0.419</td>
</tr>
<tr>
<td></td>
<td>5%</td>
<td>0.880</td>
<td>0.890</td>
<td>0.914</td>
<td>0.911</td>
<td>0.802</td>
<td>0.612</td>
<td>0.454</td>
</tr>
<tr>
<td></td>
<td>1%</td>
<td>0.913</td>
<td>0.927</td>
<td>0.949</td>
<td>0.943</td>
<td>0.869</td>
<td>0.691</td>
<td>0.515</td>
</tr>
<tr>
<td>30</td>
<td></td>
<td>0.787</td>
<td>0.813</td>
<td>0.865</td>
<td>0.889</td>
<td>0.796</td>
<td>0.616</td>
<td>0.449</td>
</tr>
<tr>
<td></td>
<td>5%</td>
<td>0.803</td>
<td>0.827</td>
<td>0.877</td>
<td>0.901</td>
<td>0.826</td>
<td>0.657</td>
<td>0.472</td>
</tr>
<tr>
<td></td>
<td>1%</td>
<td>0.829</td>
<td>0.852</td>
<td>0.897</td>
<td>0.925</td>
<td>0.890</td>
<td>0.722</td>
<td>0.542</td>
</tr>
<tr>
<td>50</td>
<td></td>
<td>0.760</td>
<td>0.790</td>
<td>0.851</td>
<td>0.889</td>
<td>0.826</td>
<td>0.638</td>
<td>0.468</td>
</tr>
<tr>
<td></td>
<td>5%</td>
<td>0.774</td>
<td>0.802</td>
<td>0.860</td>
<td>0.898</td>
<td>0.860</td>
<td>0.671</td>
<td>0.494</td>
</tr>
<tr>
<td></td>
<td>1%</td>
<td>0.793</td>
<td>0.824</td>
<td>0.878</td>
<td>0.914</td>
<td>0.909</td>
<td>0.729</td>
<td>0.548</td>
</tr>
<tr>
<td>70</td>
<td></td>
<td>0.743</td>
<td>0.778</td>
<td>0.845</td>
<td>0.891</td>
<td>0.828</td>
<td>0.654</td>
<td>0.488</td>
</tr>
<tr>
<td></td>
<td>5%</td>
<td>0.754</td>
<td>0.787</td>
<td>0.855</td>
<td>0.899</td>
<td>0.858</td>
<td>0.689</td>
<td>0.513</td>
</tr>
<tr>
<td></td>
<td>1%</td>
<td>0.774</td>
<td>0.809</td>
<td>0.867</td>
<td>0.914</td>
<td>0.906</td>
<td>0.756</td>
<td>0.560</td>
</tr>
<tr>
<td>90</td>
<td></td>
<td>0.735</td>
<td>0.771</td>
<td>0.841</td>
<td>0.892</td>
<td>0.838</td>
<td>0.667</td>
<td>0.492</td>
</tr>
<tr>
<td></td>
<td>5%</td>
<td>0.744</td>
<td>0.781</td>
<td>0.848</td>
<td>0.899</td>
<td>0.865</td>
<td>0.700</td>
<td>0.517</td>
</tr>
<tr>
<td></td>
<td>1%</td>
<td>0.767</td>
<td>0.802</td>
<td>0.861</td>
<td>0.910</td>
<td>0.906</td>
<td>0.759</td>
<td>0.574</td>
</tr>
<tr>
<td>100</td>
<td></td>
<td>0.734</td>
<td>0.770</td>
<td>0.840</td>
<td>0.890</td>
<td>0.846</td>
<td>0.674</td>
<td>0.500</td>
</tr>
<tr>
<td></td>
<td>5%</td>
<td>0.744</td>
<td>0.779</td>
<td>0.846</td>
<td>0.897</td>
<td>0.871</td>
<td>0.706</td>
<td>0.527</td>
</tr>
<tr>
<td></td>
<td>1%</td>
<td>0.761</td>
<td>0.794</td>
<td>0.862</td>
<td>0.908</td>
<td>0.913</td>
<td>0.765</td>
<td>0.583</td>
</tr>
<tr>
<td>150</td>
<td></td>
<td>0.720</td>
<td>0.760</td>
<td>0.833</td>
<td>0.892</td>
<td>0.857</td>
<td>0.682</td>
<td>0.510</td>
</tr>
<tr>
<td></td>
<td>5%</td>
<td>0.729</td>
<td>0.768</td>
<td>0.839</td>
<td>0.897</td>
<td>0.880</td>
<td>0.708</td>
<td>0.536</td>
</tr>
<tr>
<td></td>
<td>1%</td>
<td>0.746</td>
<td>0.779</td>
<td>0.852</td>
<td>0.908</td>
<td>0.926</td>
<td>0.767</td>
<td>0.595</td>
</tr>
</tbody>
</table>
Fig. 1: Performance of the statistics for $n = 50$ and $\rho = 0.90$
Fig. 2: Performance of the statistics for $n = 50$ and $\rho = 0.975$
On discordance test for the wrapped normal data

Fig. 3: Performance of the statistics for $n = 20$ and $\rho = 0.90$

Table 5:
Kuantan Wind Direction Data

<table>
<thead>
<tr>
<th>Year</th>
<th>Mean Surface Wind Direction (Radian)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1999</td>
<td>0.28707</td>
</tr>
<tr>
<td>2000</td>
<td>1.46071</td>
</tr>
<tr>
<td>2001</td>
<td>0.87509</td>
</tr>
<tr>
<td>2002</td>
<td>1.64563</td>
</tr>
<tr>
<td>2003</td>
<td>1.56786</td>
</tr>
<tr>
<td>2004</td>
<td>1.33478</td>
</tr>
<tr>
<td>2005</td>
<td>1.80266</td>
</tr>
<tr>
<td>2006</td>
<td>2.15736</td>
</tr>
<tr>
<td>2007</td>
<td>1.73430</td>
</tr>
<tr>
<td>2008</td>
<td>1.67275</td>
</tr>
</tbody>
</table>
Table 6: Descriptive Statistics

<table>
<thead>
<tr>
<th>Variable</th>
<th>Angles</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean Vector (µ)</td>
<td>84.65°</td>
</tr>
<tr>
<td>Length of Mean Vector (r)</td>
<td>0.88</td>
</tr>
<tr>
<td>Concentration</td>
<td>3.33</td>
</tr>
<tr>
<td>Circular Variance</td>
<td>0.12</td>
</tr>
<tr>
<td>Circular Standard Deviation</td>
<td>28.45°</td>
</tr>
<tr>
<td>Standard Error of Mean</td>
<td>10.57°</td>
</tr>
</tbody>
</table>

Fig. 4: Circular Plot of Kuantan Wind Data

Table 7: Result based on C, M, D and A statistics

<table>
<thead>
<tr>
<th>Test</th>
<th>Test value</th>
<th>Cut-off point</th>
<th>Decision</th>
</tr>
</thead>
<tbody>
<tr>
<td>C</td>
<td>0.07</td>
<td>0.08</td>
<td>Not an outlier</td>
</tr>
<tr>
<td>M</td>
<td>0.59</td>
<td>0.64</td>
<td>Not an outlier</td>
</tr>
<tr>
<td>D</td>
<td>0.13</td>
<td>0.19</td>
<td>Not an outlier</td>
</tr>
<tr>
<td>A</td>
<td>0.60</td>
<td>0.61</td>
<td>Not an outlier</td>
</tr>
</tbody>
</table>
QUANTIFYING THE EFFECTS OF TURBULENT ATMOSPHERE ON SMALL SATELLITES AT PAKISTAN IONOSPHERIC ENVIRONMENT

M. Ayub Khan Yousuf Zai\(^1\) and Khusro Mian\(^2\)

\(^1\) Solar Terrestrial and Atmospheric Research Center, Department of Applied Physics, University of Karachi, Karachi, Pakistan. Email: ayubzai@yahoo.com
\(^2\) Institute of Space and Planetary Astrophysics (ISPA) University of Karachi, Karachi, Pakistan

ABSTRACT

The ionosphere is one of the constituents of thermosphere of the upper atmosphere. Upper atmosphere as whole is an open system that is regarded as a non linear system and that seems to be complex. Therefore, a non linear mathematical model is plausible to explain phenomena of ionospheric turbulence. In this communication we have paid our attention in the effects of turbulent atmosphere on small satellites at Pakistan ionospheric environment. The dominance of turbulence on small satellites passing through ionosphere has been documented. Multi regression analysis has been utilized to come across the involvement of perturbation in ionospheric region. Goodness of fit test and non parametric test are used for normality and randomness in data respectively. These results might be useful in parameterization of Astrophysical plasma.

KEYWORDS

Thermosphere; open system; turbulent atmosphere; small satellites; ionospheric environment; multi regression; nonlinear; perturbation

1. INTRODUCTION

Turbulence means unstable and disorderly movements and really the ultimate result of current of air, water and moving with circular motion against the main current, it is rotatory motion state of different size and shape. Ionospheric physics as contemporary science expresses different views such as solar terrestrial mechanisms, earth’s environmental physics radio communication, including prediction and forecasting, estimation and monitoring the turbulence, analyzing the effects of ionospheric layers on the radio wave propagation, natural electromagnetic emission, magnetic pulsations and other phenomena. Ionospheric region for active experiments as natural plasma, often nonlinear and non thermal with radio waves, powerful produce modifications in the ionosphere in the form of enhanced temperatures and plasma instabilities.

As a constituent of upper atmosphere plasma provides new physical conceptions and space weather comprise the variability that is regarded as short term process of the ionospheric region of the major irregularities in the Es and F\(_2\) layers. The process of creating these irregularities begins with share forces between two masses. This force creates large eddies spawn small and smaller eddies, transferring in the transformation of inertial range.
The large electric fields generated during thunderstorms can adversely affect ionosphere and possible relations between solar and magnetic disturbances and weather have been found but physical interpretation is still not recognized. The effect of turbulence is to scatter a small portion of incident energy is redirected and provides a week signal outside the incident beam tropo-scatter is always present to some extent and can provide communication beyond the horizon paths. Tropo-scatter can also cause interference to communication system operating well beyond the horizon [Smirnov B M et al., 1977; Chandrasekhar S et al., 1960]

In this study we intend to enumerate the effects of variability in the form of turbulence on the small satellites for lower earth observations at Pakistan ionospheric environment at mid latitudes, diurnal fluctuations in the ionic concentration of the F_2 has been computed that can cause geomagnetic disturbances, waves and tides from lower atmosphere and changes in the solar radiations. Ionospheric perturbations can be enumerated and this life time fluctuation could be estimated and we can investigate that monitoring of this important region of our globe is as critical for progress in the future as it was in the past [Dolukhanov M et al., 1971; Cairnset R A et al., 1985].

2. IONOSPHEREIC LAYERS

The altitude of the ionospheric layers are situated about (90 – 1000) Km, there are different D, E and F layers. They have different behaviors each layer. Since there are two anomalies in phase in the northern hemisphere, variations in the critical frequency there are particularly pronounced. In the southern hemisphere, they are in opposition, which give rise to a biannual anomaly, due to this critical geographical position and the large positive correlation between the plasma turbulence and yearly variation of time in the ionospheric region [Leontovich M et al., 1965; Galee A A and Sudan R N et al., 1984; Chen M C et al., 1929; Barclay L et al., 2003]

3. COMPUTATION OF TURBULENCE FLUX

We have computed Turbulence flux upper atmospheric Ionospheric layer at Pakistan air space, by using perturbed concentration, and perturbed parcel velocity of ionospheric layer. Turbulent flow varies randomly with time at a location as shown in Fig.1 thus turbulent flow is unsteady. Nearly all flows in the atmosphere are turbulent. Sub grid-scale effects are estimated by substituting decomposed variables into an equation. The product $\frac{U}{\nu} \cdot N$ represents the transport of Kinematics Turbulent Flux:

$$\frac{U}{\nu} \cdot N = \frac{U'_E N'_E + U'_F N'_F}{2}$$

$U'_E$ An Instantaneous and Local Perturbation Scalar Velocity content

$U'_F$ An Instantaneous and Local Perturbation Scalar Velocity content

$N'_E$ An Instantaneous and Local Perturbation Plasma Concentration content

$N'_F$ An Instantaneous and Local Perturbation Plasma Concentration content

[Mark Z. Jacobson et al, 1999]
4. ANALYSIS OF THE DATA

Multi-regression as nonlinear data analysis is very genuine numerical techniques in mathematical models, which analyzing the behavior of turbulence at upper atmosphere such as effects of turbulent atmosphere on small satellites at Pakistan ionospheric environment. The ionospheric data was recorded by Digisonde installed at SUPARCO Karachi station which is monitoring ionospheric data as a parameter of critical frequency in mega hertz (MHz) and virtual height in Kilometer (Km) which is uncomplicated to additional investigation. We have hourly data for our investigation.

We are using the turbulence as dependent variable and electron concentration as the independent variable then we could obtain a curve model. Upper atmosphere system, where there have been always complexity and natural phenomena are open system, that exhibit randomness. Small sample of the objectives in quantifying open system is to explain turbulence and perturbation in parametric nonlinear process study of turbulence in upper atmospheric plasma at Pakistan air space. Turbulence effects are quantified by non linear mathematical methods as given.

Unification between X and Y

\[ X \equiv \text{Electron Concentration} \times 10^{12} \ [m^{-3}] \] and

\[ Y \equiv \text{Plasma Turbulence} \times 10^{12} \ [m^{-2} s^{-1}] \]

Coefficient of Determination = 0.906
Square Coefficient of Determination = 0.821
P Value = 0.00

(A) MODEL EQUATION:

\[ Y = -0.12 + 1.57X + 0.006X^2 \] (1)

Intercept or Constant value
\[ \beta_0 = -0.120078 \]
Coefficient of X \[ \beta_1 = 1.574002 \]
Coefficient of \[ X^2 \] \[ \beta_2 = 0.005772 \]

Association among dependents and independents value

\[ F_{\text{Test}} = 831.16 \] (2)
\[ t(362) = -0.6719 \] (3)
\[ \text{Standard Error} = 0.075 \] (4)

[Holger Kantz and Thomas Schreibe et al., 2002; Daniel Yates S, David Moar S and George Cabe P Mc et al., 1996].

With the help of analysis of variance, the coefficient of determination can be calculated

\[ R^2 = \frac{SS}{\text{total SS}} \]
This indicates that 82 percent of the variation is explained by the regression model. The remaining 18 percent of the variation is itself unexplained. It shows that a good forecast of the value of \( x \) is possible when the previous value of \( x \) is known. Quantifying the association between electron concentration and plasma turbulence are significance result in this system as shows in the table 1 [Holger Kantz and Thomas Schreibe et al., 2002; Daniel Yates S, David Moar S and George Cabe P Mc et al., 1996]

(B) Validity of the Constructed model

A comparison of the observed with the predicted value comfortably establishes the validity of the constructed model. We can use residual analysis to construct the original time series. In this study, subtraction of the reconstruct ionospheric turbulence data from the original computed record yields are the residual component of the time series. The residual records are primary interest which is justified our computed model. The residual analysis is shown in figures 2,3 and 4 demonstrating that the constructed model is reasonably adequate. There are some residual analysis use for real time series approaches Predicted against Observed Values in Figure 2, Observed Values against Residuals in Figure 3 and Predicted against Residual Scores in Figure 4 [Emery W J et al., 2004]

(C) GOODNESS OF FITTING TEST

It is maximum absolute deviation of expected relative frequency \( F_e \) and the observed relative frequency \( F_0 \). It is denoted by \( D \). The data computed as a theoretical distribution, the Kolmogorov-Smirnov Test was used. Figure 5 depicts the distribution of maximum turbulence of ionosphere to apprises the behaviour of the data that emerge to come from normal. It is one tailed test for level of significance \( \alpha \).Kolmogorov-Smirnov Test for turbulence \( (x \times 10^{12}) \) in ionospheric layers at Pakistan air space for goodness of fitting K-S, \( D = 0.07323, p < 0.01 \) in table 2 and figure 4

\[
i) \quad D = |F_0 - F_e| = 0.07
\]

\[
ii) \quad \text{The tabulated value of } D \text{ for } n = 6 \text{ and } \alpha = 0.01 \text{ is } 0.07323
\]

Science, the table value of \( D \) is 0.07323 is greater than the calculated value of 0.07, so we accept the null hypothesis the distribution follows a normal distribution. [Arora P N, Sumeet Arora et al., 2000]

(D) RANDOMNESS TEST

Randomness in data is tested by non parametric test. The values falling above and below the median of the sample, the resulting series of lower and upper value between median can then be tested for randomness on the basic of the total number of runs of series of lower and upper value between median of turbulence \( (x \times 10^{12}) \) in ionospheric layers at Pakistan air space,

\[
N_1 \equiv \text{Numbers of greater than median},
N_2 \equiv \text{Numbers of less than median}
\]

Null Hypothesis: \( H_0 \): the sample is random
Alternative Hypothesis: \( H_1 \): the sample is not random
Level of significance \( \alpha = 0.05 \),
Test statistics: Median value of Turbulence ionospheric layers = 2.343106, \( N_1 = 182 \), \( N_2 = 182 \), \( V = 180 \) (any random value less than \( N_1 = 182 \) and \( N_2 = 182 \))

\[
\mu = \frac{2N_1N_2}{N_1 + N_2} + 1 = 182.5 
\] (6)

\[
\sigma^2 = \frac{2N_1N_2(2N_1N_2 - N_1 - N_2)}{(N_1 + N_2)^2(N_1 + N_2 - 1)} 
\] (7)

\[= 90.5 \]

\[
Z = \frac{V - \mu}{\sigma} = -0.3 
\] (8)

\[|Z| = 0.3 \] (9)

Critical value: the value of \(|Z_\alpha|\) for \( \alpha = 0.05 \) from the table is \(|Z_\alpha| = 1.96 \)

Decision: since the calculated value of \(|Z| < \text{Tabulated value of } |Z_\alpha|\) as \(0.3 < 1.96\),

Null Hypothesis: \( H_0 \) is accepted the sample is random.

[Arora P N, Sumeet Arora et al., 2000]

5. COMMENTS AND CONCLUSION

In this paper we presented an idea to find out the hidden prediction and different parameters quantifying the effects of turbulent atmosphere on small satellites at Pakistan ionospheric environment. The more precise results obtained using hypothesis test for randomness and normality in data. These results might be useful in astrophysical plasma especially in upper atmospheric layers that contain plasma.

ACKNOWLEDGEMENTS

The authors are grateful to the officials of Space and Upper Atmosphere Research Commission (SUPARCO) who provides us the Ionospheric data in the form of critical frequency and other parameters recorded by Digital Ionospheric Sounder DGS-256.

Also the local hospitality offered by the organizers of this conference is greatly appreciated.
**Fig. 1.** Temporal behaviour of Computed Turbulence for the year 1989

**Fig. 2:** Predicted value vs. observed value of fitting model for turbulence and electron concentration at ionospheric F$_2$ layer.

**Fig. 3:** Observed values vs. residuals value of fitting model for turbulence and electron concentration at ionospheric F$_2$ layer.
Fig. 4: Predicted value vs. residual values of fitting model for turbulence and electron concentration at ionospheric F$_2$ layer.

Fig. 5: Kolmogorov–Smirnov test Histogram values of fitting model for turbulence and electron concentration at ionospheric F$_2$ layer.

Table 1:

<table>
<thead>
<tr>
<th></th>
<th>SS</th>
<th>D f</th>
<th>MS</th>
<th>F</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reg.</td>
<td>176.3</td>
<td>2</td>
<td>88.16</td>
<td>831</td>
<td>0.0</td>
</tr>
<tr>
<td>Res.</td>
<td>38.39</td>
<td>362</td>
<td>0.10</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sum</td>
<td>214.7</td>
<td>364</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 2:
Frequency distribution table for goodness fitting test turbulence model, D= |F₀ - Fₑ|

<table>
<thead>
<tr>
<th>C-I</th>
<th>F</th>
<th>F₀</th>
<th>Fₑ</th>
<th>D</th>
</tr>
</thead>
<tbody>
<tr>
<td>0-1</td>
<td>4</td>
<td>0.11</td>
<td>0.04</td>
<td>0.07</td>
</tr>
<tr>
<td>1-2</td>
<td>103</td>
<td>0.29</td>
<td>0.28</td>
<td>0.01</td>
</tr>
<tr>
<td>2-3</td>
<td>189</td>
<td>0.81</td>
<td>0.76</td>
<td>0.05</td>
</tr>
<tr>
<td>3-4</td>
<td>55</td>
<td>0.96</td>
<td>0.98</td>
<td>0.02</td>
</tr>
<tr>
<td>4-5</td>
<td>9</td>
<td>0.98</td>
<td>0.99</td>
<td>0.01</td>
</tr>
<tr>
<td>5-6</td>
<td>5</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>

365

REFERENCES
    84 Theobalts Road London WCIX8RR UK.
ON THE ROLE OF THE MULTIVARIATE FATIGUE LIFE DISTRIBUTION
IN MODELLING MULTI-SITE FATIGUE

Saleha Naghmi Habibullah¹ and Sam C. Saunders²
¹ Kinnaird College for Women, Lahore, Pakistan
Email: salehahabibullah@hotmail.com
² Prof. Emer. Washington State University, USA

ABSTRACT

This paper reviews the role of the univariate fatigue life distribution in modeling the life-length of a component undergoing incremental cumulative damage (due to wear, abrasion, etc.) and extends the ideas to structures of multiple components. It is emphasized that, because of its connection with the physical realities of the cumulative damage processes experienced by industrial products, it seems to carry great potential for modeling fatigue lives of physically related components. However, the worth of a newly derived model being demonstrable only in a practical situation, its potential and applicability needs to be explored in an industrial setting.

KEYWORDS AND PHRASES

Fatigue Life Distribution; Self-inversion; Late-life Plateau; Bivariate Fatigue Life Distribution; Multivariate Fatigue Life Distribution.

1. INTRODUCTION

The Fatigue Life distribution was published by Birnbaum and Saunders (1969a) more than forty years ago. It was a model for life-length of aluminium airframe components which would eventually fail from fatigue during the repetition of cyclic stresses while in service. The distribution was initially applied to calculate safe-life in durability studies for the Boeing Commercial Airplane Division in Renton during the 1960’s. Since then it has been successfully applied to various types of service lives limited by cumulative damage, viz., wear, abrasion, galling, rust and nuclear or UV-radiation. Also, it has been used in mortality studies and in environmental contamination exceedance investigations as well as in other applications all over the world.

But complex systems, structures or machines often contain subsets of identical, often physically adjacent components, which are exposed to the same stresses during their duty-cycle replications in service, as well as the same ambient conditions including humidity, temperature and exposure to radiation. In this paper, we discuss a current investigation of a multivariate probability model for dealing with associated subsets of components sustaining the same damaging conditions. The goal is to develop a multivariate probability distribution representing the related life-lengths of components undergoing identical stochastic wear-processes.

It seems reasonable to assume that very similar conditions may result in highly correlated times of failures when accumulated damage is virtually identical causing the
same reductions in the yield stress. An application might be the multi-site fatigue, such as is encountered by components in nuclear generating plants, deep sea platforms or in airframes (as well as in other structures).

2. RE-VISITING THE UNIVARIATE FATIGUE LIFE DISTRIBUTION

We begin with the simplest situation where we are dealing with incremental damage occurring during duty cycles of equal duration such as successive explosions in the cylinder of an internal combustion engine.

2.1 The Case of Failure from Cumulative Damage

Let \( X_i > 0 \) represent random non-negative (i.e. irreversible) incremental damage during the \( i^{th} \) duty cycle which is of duration \( \epsilon > 0 \), and they are \( i.i.d \) variates for \( i = 1, 2, \ldots \). (We assume here, for mathematical simplicity, the duty cycles are all of the same duration as in the ignition explosions in an internal combustion engine.)

The mean and variance of incremental damage are:

\[
E[X_i] = \epsilon \mu \quad \text{Var}[X_i] = \epsilon \sigma^2 \quad \text{for all } i = 1, 2, \ldots
\]

The total duration of time that will have elapsed by the end of the \( n^{th} \) duty cycle is \( n \epsilon \). We call this duration of time the service interval.

Let the cumulative damage after \( n \) duty cycles be denoted by \( D_n = \sum_{i=1}^{n} X_i \). We note that:

\[
E[D_n] = n \epsilon \mu \quad \text{Var}[D_n] = n \epsilon \sigma^2
\]

Let \( \omega \) be the critical damage level which when exceeded becomes the "stopping event" either by causing failure or, in some way, the cessation of further duty cycles. Also, let \( N \) be the random number of duty cycles necessary to first exceed this critical limit \( \omega \). Then the random service period required until failure of the component will be \( T = N \epsilon \). We then have, by duality, an equality of probability between the events

\[
\Pr[T > n \epsilon] = \Pr[N > n] = \Pr[D_n \leq \omega] = \Pr[\frac{D_n - n \epsilon \mu}{\sqrt{n \epsilon \sigma}} \leq \frac{\omega - n \epsilon \mu}{\sqrt{n \epsilon \sigma}}] = \Pr[\frac{n \epsilon \mu - D_n}{\sqrt{n \epsilon \sigma}} \geq \frac{\omega - n \epsilon \mu}{\sqrt{n \epsilon \sigma}}]
\]

Since \( T = N \epsilon \), hence, by setting

\[
t = n \epsilon, \quad \text{and} \quad Z_n = \frac{n \epsilon \mu - D_n}{\sigma \sqrt{n \epsilon}}
\]

we obtain

\[
\Pr[T > t] = \Pr[\frac{n \epsilon \mu - D_n}{\sqrt{n \epsilon \sigma}} \geq \frac{n \epsilon \mu - \omega}{\sqrt{n \epsilon \sigma}}] = \Pr \left[ Z_n \geq \frac{\omega \mu}{\sigma} \left( \sqrt{\frac{\mu}{\omega \mu} - \sqrt{\frac{\mu}{\omega \mu}}} \right) \right]
\]

Now, by letting \( \alpha = \frac{\sigma}{\sqrt{\omega \mu}} \), \( \beta = \frac{\omega}{\mu} \) and defining

\[
\xi(x) = 2 \sinh[\ln \sqrt{x}] = x^{1/2} - x^{-1/2} \quad \forall \quad x > 0
\]

we obtain

\[
\Pr[T \leq t] = \Pr \left[ Z_n < \frac{1}{\alpha} \left( \sqrt{\frac{\mu}{\beta}} - \sqrt{\frac{\beta}{\mu}} \right) \right] = \Pr \left[ Z_n < \frac{1}{\alpha} \xi \left( \frac{t}{\epsilon} \right) \right]
\]
Now, if we let \( n \to \infty \) and \( \epsilon \to 0 \) with \( t = n \epsilon \) fixed, then, by the CLT we know that \( Z_n \sim \mathcal{N}(0,1) \) as \( n \to \infty \). So we obtain, in the limit,

\[
F(t) := \Pr[T \leq t] = \Phi[\frac{t}{\alpha \beta}] \quad \forall \ t > 0; \ \alpha, \beta > 0,
\]

where

\[
\Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-t^2/2} \, dt \quad \text{with} \quad \Phi' = \varphi.
\]

This is the fatigue-life distribution \( \mathcal{FL}(\alpha, \beta) \) with only two parameters: \( \beta \), the median (scale parameter) and \( \alpha \), the shape parameter determining the coefficient of variation. The mean, variance and coefficient of variation of the (time-to-failure), fatigue-life \( T \), are given by

\[
E[T] = \beta(1 + \frac{\alpha^2}{2}), \quad \text{Var}[T] = (\alpha \beta)^2[1 + \frac{5\alpha^2}{4}],
\]

and

\[
\text{cvf}[T] = \frac{\sqrt{1 + \frac{5\alpha^2}{4}}}{1 + \frac{\alpha^2}{2}}.
\]

**Remark:**

In this elegant derivation by Marshall and Olkin (2007), all that is required is that the Central Limit Theorem hold. The incremental damage variates need not be identically distributed or even all independent as long as they are non-negative and uniformly and asymptotically negligible (Loéve’s phrase). (Of course the means and variances need to be Cesaro summable, i.e.

\[
\frac{1}{n} \sum_{i=1}^{n} \mu_i \to \mu \quad \text{as} \quad n \to \infty, \quad \text{and} \quad \frac{1}{n} \sum_{i=1}^{n} \sigma_i^2 \to \sigma^2 \quad \text{as} \quad n \to \infty.
\]

Moreover the random stopping number need not be cumulative damage but does have to be a function of the history, namely, the event \([n = n]\) is a function only of \(X_1, X_2, \ldots, X_n\).

Many other cumulative damage situations can also be described in terms of stochastic processes. For example, the damage at time \( t \) may be given by \( D(t) = \sum_{i=1}^{N} X_i \) where \( X_i > 0 \) are iid and \( N \) has the Poisson distribution, namely

\[
P[N = k] = \frac{e^{-\lambda}(\lambda t)^k}{k!} \quad k = 0, 1, \ldots
\]

Now, for any value of \( t > 0 \), let \( \lambda \uparrow \infty \) and we obtain the \( \mathcal{FL} \)-distribution in the limit.

If we consider the situation when the component wear is negligible during service but "start-up" is the chief cause of damage (like an incandescent light bulb) so that the lengths of the duty cycles do not matter but only their number is important, then we must determine the distribution of \( N = \text{the number of cycles} \) at failure and it too may have a fatigue-life distribution when the cycle-number for failure is sufficiently large and the range of damage for the incremental variates is not too concentrated.
In other words: In almost any practical case where ‘failure’ is the result of the gradual accumulation of a large number of small non-negative incremental steps, the random time until failure can very likely be modelled by the Fatigue-Life distribution.

(NB: The random waiting time $T$ until the critical level of damage is exceeded has a fatigue-life distribution only because no ‘healing’ can occur during damage accrual; if the increments can be positive or negative, then the limiting distribution is not the Fatigue Life distribution.)

Of course, no one can prevent Nature from producing a situation when we must consider both number of duty cycles and the amount of wear incurred during service in each duty cycle, which itself is of random duration. Then a bivariate distribution, which will contain both discrete and continuous variates, must be analysed. Thus, it is still an open problem to determine the asymptotic distribution for the expression

$$\Pr \left[ N = n, \sum_{i=1}^{n} D_i(T_i) < \omega \right]$$

where $D_i(T_i)$ is the damage during the $i^{th}$ duty cycle which itself is of random duration $T_i$ for $i = 1, \ldots, n$.

2.2 Some Other First Passage Time Distributions

In 1915 Schrödinger (1915) found the density for the first passage time of a particle to travel a distance $\ell$ when moving at an expected velocity of $\nu$ under Brownian motion. He used his knowledge of the heat-diffusion equation plus the principles of reflection and diffusion to obtain his “Schrödinger" density

$$f_S(t) = \frac{\ell^{3/2}}{\sigma \sqrt{2\pi}} \exp\left\{ -\frac{(\ell - \nu t)^2}{2\sigma^2} \right\} \quad \forall \; t > 0.$$  

But it went without notice, presumably because it was thought to apply only to Brownian Motion.

In 1944 Wald (1944) showed that, for a sequence $\{X_i\}_{i=1}^{\infty}$ where the $X_i$ are iid with $E[X_i] = \nu > 0$ and $\text{Var}[X_i] = \sigma^2$, the event first exceedence of the critical amount $\omega$, in gain or loss in $n$ plays

$$[N = n] = \left[ |X_1| \leq \omega, \ldots, |\sum_{i=1}^{n-1} X_i| \leq \omega, |\sum_{i=1}^{n} X_i| \geq \omega \right]$$

has asymptotically the Wald distribution. The asymptotic distribution $F_W$ derived by Wald is

$$\Pr[N \leq n] \approx F_W(n) = \int_{0}^{y} \frac{x^{-3/2}e^{-x^2}}{\sigma \sqrt{2\pi}} \exp\left\{ \frac{-1}{2\sigma^2} (x - x^{-1})^2 \right\} \, dx \quad \forall \; y > 0.$$  

But the density couldn’t be integrated in closed form. The same distribution was later investigated by Zigangirov (1962) and Shuster (1968).
2.3 First Passage Variate N in Number of Observations:

Let $\mathcal{N}$ denote the set of natural numbers and let

$$N = \min\{n \in \mathcal{N} : \sum_{i=1}^{n} X_i > \omega \text{ or } \sum_{i=1}^{n} X_i < -\omega\}$$

where $X_i$ can be positive or negative.

The distribution must be evaluated by considering

$$\Pr[N > n] = \Pr[|X_1| < \omega, |X_1 + X_2| < \omega, \ldots, |\sum_{i=1}^{n} X_i| < \omega]$$

This chain of arguments was presented by Wald (1947) in his treatise on Sequential Analysis. As well, it is suitable for modeling a situation involving continued wagering until gaining or losing $\omega$ dollars. As such, the Wald distribution may be appropriate for a variety of situations such as investment theory but not for non-negative incremental damage.

In 1957, Tweedie (1957) inverted the characteristic function of the Gaussian distribution to obtain the Inverse Gaussian distribution. By reparameterization one sees the equality of Schrödinger’s, Wald’s and Inverse-Gaussian distributions. Let $F_S$, $F_W$ and $F_{IG}$ be the cdf’s of the three distributions. Then, through reparametrization, they all can be expressed as

$$F(t) = \Pr[T \leq t] = \Phi\left[\frac{1}{\alpha} \xi\left(\frac{t}{\beta}\right)\right] + e^{2\alpha^2} \Phi\left[\frac{-1}{\alpha} \psi\left(\frac{t}{\beta}\right)\right] \quad \text{for } t > 0, \alpha, \beta > 0$$

where

$$ET = \beta, \quad \text{Var}[T] = (\alpha\beta)^2$$

and, for notational convenience here and later, we write

$$\xi(x) = x^{1/2} - x^{-1/2}, \quad \psi(x) = x^{1/2} + x^{-1/2} \quad \forall \ x > 0.$$

2.4 Tweedie’s RIG Distribution

The RIG variate, say $T$, is defined by

$$\frac{T}{\beta} \sim \mathcal{IG}(\alpha, \beta) \quad \text{i f f} \quad \frac{\beta}{T} \sim \mathcal{IG}(\alpha, \beta)$$

The distribution is given by

$$\Pr[T < t] = \Phi\left[\frac{1}{\alpha} \xi\left(\frac{t}{\beta}\right)\right] - e^{2\alpha^2} \Phi\left[\frac{-1}{\alpha} \psi\left(\frac{t}{\beta}\right)\right]$$

and we have

$$ET = \beta(1 + \alpha^2), \quad \text{Var}[T] = (\alpha\beta)^2(1 + 2\alpha^2)$$

2.5 Applicability of the Fatigue Life Distribution:

An important question is: when is the use of the Fatigue-life variable advisable for describing (predicting) life? The answer lies in the realization that among life distributions the asymptotic behavior of the hazard rate must match the empirical facts. Among the class of first passage time distributions, only the Fatigue Life distribution has the long-life plateau effect. The concept of the "plateau" is explained subsequently.
2.6 The Late-Life Plateau

The traditional bath-tub-shaped hazard-rate is given in Figure 1 whereas, a “whole-life” hazard-rate containing the plateau effect is given in Figure 2.

Fig. 1: Traditional Bathtub-Shaped Hazard Rate

Fig. 2: The Fatigue Life Late-Life Plateau

That is, the hazard rate tends to a positive constant as \( t \to \infty \).

As far as biological life is concerned, a number of papers have challenged the long-standing notion that mortality rates will keep on increasing (monotonically) with age. Greenwood and Irwin (1939) comment that the increase of mortality rate with age advances at a slackening rate and suggest the possibility that “with advancing age the rate of mortality asymptotes to a finite value”. They also find the same regularities as in humans as is evident from their remark "Some tests of the ultimate mortalities in non-human experience were not unfavorable."

In the 1960’s Vallodi Weibull conducted fatigue experiments on 1000 samples of metal wire. He was surprised to find that the observations couldn’t be fitted with a single Weibull distribution because of the deceleration of the hazard rate at late life. (See Weibull, 1967).

In addition to a demonstration of mortality levelling-off at advanced ages for invertebrates (including fruit flies and house flies)and rodents, Economos (1979) describes mortality levelling off for several manufactured products. The author comments that mortality kinetics of various animal species and failure kinetics of industrial components and materials are at variance with Gompertz’s law or law of exponentially increasing force of mortality, and concludes that after a certain species-
characteristic age, force of mortality and probability of death cease to increase exponentially with age, and remain constant at a high level on the average for the remainder of the life span.

The phenomenon of mortality deceleration at late ages seems to have caught the greater attention of researchers due to the publication of book by Gavrilov and Gavrilova (1991) "The Biology of Life Span: A Quantitative Approach" (the English translation of the original 1986 Russian original) as well as the publication of the paper by careyetal92 and by curtsiertaletal92. gavgav91 dismiss the widely publicized work of hayflick65, who argues for the existence of a maximum lifespan based on the identification of a finite limit to cell division, the authors elaborate their own theory, called "the limited reliability of organisms."

According to Olshansky (1992), the premise is simple and intuitively appealing: the decline of living biological organisms and of nonliving mechanical devices is remarkably similar; their comparable rate of breakdown can be proven mathematically; in both cases there is high rate of failure early on, followed successively by relatively low and stable failure rates, exponentially increasing failure rates, and finally a deceleration of that increase near the end of life. The aging of living and nonliving systems is therefore believed to result from "a cascade of dependent failures’ which occurs when one of the organism’s systems randomly fails". Since living organisms and inanimate mechanical devices are saturated with defects from the outset, individual systems fail and death (or nonfunctioning) occurs as a result of stochastic events operating on existing defects.

Carey et al. (1992) report that mortality rates levelled off and decreased at older ages in a population of 1.2 million medflies maintained in cages containing 7200 each, and in a group of approximately 48,000 adult medflies maintained in solitary confinement. Curtsinger et al. (1992) report that analysis of large cohorts of ten genotypes of Drosophila melanogaster raised under conditions that favored extended survival showed that mortality rate leveled off at older ages.

Kannisto et al. (1994) present a detailed and comprehensive account of the reduction in mortality at advanced ages basing their arguments on several decades of evidence from 27 countries.

Rauser et al. (2006) comment that late life is a distinct phase of life characterized by a cessation in the deterioration of survivorship and that several theories have been proposed to explain non-aging at late ages, specifically with regards to late-life mortality-rate plateaus. They opine that all such theories must be compatible with formal evolutionary theory and experimental findings, and go on to developing a critique of theories of late life based on evolutionary biology.

But the late-life plateaus are not restricted to biological life only. According to the IEEE Spectrum (Nov.2004), "The fourth epic in life is another one that we and our machines share....."; "There have been numerous attempts to explain the biology behind this but since this is found in man-made stuff, such as steel, relays, thermal insulation, perhaps Reliability Theory may provide a better answer."

Lai (2010) proposes some simple generalizations of the Erlang distribution and examines the mortality rate functions and their shapes which inform regarding the
manner in which populations age under these models. The author asserts that, with suitable selections of parameter values, we can achieve all the common monotonic or non-monotonic hazard rate shapes that occur in reliability and survival analysis; some of these functions have an S shape which is suitable for the lifetime random variables that exhibit a late life deceleration phenomenon.

Bebbington et al. (2011) derive a new distribution based on the Strehler-Mildvan theory of aging and apply it to mortality data from the 1892 cohort of Swedish women and on Mediterranean fruit flies. For the distribution fitted to the Swedish mortality Data, the force of mortality begins to decelerate at 89 years and begins to decrease at 108 years, where the data extend to 110 years. The authors opine that, excepting infant mortality, the newly derived mortality distribution captures many possible shapes of late-life mortality.

**MODELING THE LATE-LIFE PLATEAU**

As far as the use of the lognormal distribution is concerned, for this distribution, the hazard rate ultimately goes to zero, and the mean residual life increases. This behavior contradicts the plateau effect discussed earlier. As far as the use of the Weibull distribution is concerned, since it is a Gumbel extreme minimum value distribution, its valid applicability is in all except late life calculations.

Why not use Wald’s distribution? Because the Wald distribution does not fit the situation which is virtually always encountered in industry, i.e. the damage increments (having stationary distributions) not being negative. (Consequently, even the life of electronic systems exhibits the "plateau" effect of the hazard rate.) Why not use Tweedie’s RIG distribution? Because, for this distribution, the hazard rate ultimately goes to zero, and the mean residual life ultimately increases.

**Theorem:**

If a life distribution $F$ is a mixture

$$F = pF_{IG} + qF_{RIG},$$

where $0 \leq p \leq 1, q = 1 - p$ and the scale and parameters match (respectively), then we have,

$$F(t) = \Phi\left[\frac{1}{\alpha} \xi(t/\beta)\right] + (p - q)e^{2\alpha^{-2}}\Phi\left[-\frac{1}{\alpha} \psi(t/\beta)\right]$$

For all $t > 0; \alpha > 0$.

For a mixture of the Wald and Tweedie-RIG distributions, we obtain an expression for its hazard rate which asymptotically tends to a constant, viz.,

**NB:** The important point is that the physical conditions of industrial damage process may not match the Wald or Tweedie’s distributions’ derivational requirements. As far as the Wald and Tweedie RIG mixture is concerned, the correction term $(p - q)e^{2\alpha^{-2}}\Phi[-\frac{1}{\alpha} \psi(t/\beta)]$ in the mixture tends to zero as $t \to \infty$, so we obtain the same asymptotic behavior as that of the Fatigue Life distribution.
SUITABILITY OF THE $\mathcal{FL}$ DISTRIBUTION FOR BIOLOGICAL AND NON-BIOLOGICAL LIFE

For machine life, the hazard rate is often devoid of early failures caused by manufacturing defects, but exhibits the late-life plateau as depicted in Figure 3.

As indicated in Section 3.2, this can easily be modeled by the fatigue life distribution.

For biological life, which exhibits both the infant mortality and the late-life plateau (resulting in a graph similar to the one contained in Figure 2), the minimum of two Fatigue Life variates yields the desired hazard rate. Let $I$ be the life variate with hazard rate $h$, and let $T_i$ be an $\mathcal{FL}$-variates with hazard rate $h_i$, such that $T_i$ is the time of death due to infant mortality (childhood diseases), and $T_2$ is the time of death due to old age from gradual weakening of the immune system. Then

$$L = \min(T_1, T_2) \Rightarrow h = h_1 + h_2$$

and, with an appropriate choice of parameters, this model yields the graph of a hazard rate exhibiting both the infant mortality and the late-life plateau.

THE SELF-INVERSION PROPERTY OF THE $\mathcal{FL}$ DISTRIBUTION

The self-inversion property of the $\mathcal{FL}$-variates viz.

$$T \sim \mathcal{FL}(\alpha, \beta) \Rightarrow T/\beta \sim \beta/T$$

effectively doubles any sample size since $T$ and $1/T$ both provide information about the parameters $\alpha, \beta$. Hence estimates made using a censored sample of “k out of n” observations can use $t_{(1)} < \cdots < t_{(k)}$ from the ordered sample as well as their reciprocals for estimation.

BIVARIATE FATIGUE LIFE DISTRIBUTION

Let us consider the case where we have two components exposed to cumulative damage. If we standaridize the $\sum_{i=1}^{n} X_{ij}$, the cumulative damage, then, in the limit $\frac{1}{\alpha_i} \zeta(T_i/\beta_i)$ becomes Gaussian for $i = 1, 2$, and each of the waiting times until damage exceedence becomes Fatigue Life distributed.
If the cumulative damage processes are related, then it seems reasonable to assume that, under mild conditions, \( \frac{1}{\alpha_i} \xi(T_i/\beta_i) \) will be jointly normal. In other words, 
\[
(T_1, T_2) \sim \mathcal{MF}_c(\alpha, \beta, \Sigma)
\]
if and only if 
\[
\left( \frac{1}{\alpha_1} \xi(T_1/\beta_1), \frac{1}{\alpha_2} \xi(T_2/\beta_2) \right) \sim \mathcal{N}(0, \Sigma)
\]
where \( \Sigma \) is a \( 2 \times 2 \) correlation matrix.

**THE BIVARIATE DENSITY FUNCTION**

The bivariate fatigue life distribution is given by

\[
f_T(t; \alpha, \beta, \Sigma) = |\Sigma|^{-1/2} \prod_{i=1}^{2} \left[ \frac{\xi(t_i/\beta_i)}{\sqrt{2\pi \alpha_i \beta_i}} \right] \times \exp \left\{ -\frac{1}{2} Q(t_1, t_2) \right\}
\]

where the quadratic form is with \( \Sigma^{-1} = (s_{i,j}) \).

**COVARIANCE & CORRELATION**

We determine the covariance between \( T_1 \) and \( T_2 \) which are both marginally fatigue-lived and with \( Z_i = \frac{1}{\alpha_i} \xi(T_i/\beta_i) \) bivariate normal. Since each of the two \( Z_i \) is standard normal, we have \( E[Z_i] = 0 \), \( E[Z_i^2] = 1 \), and, by defining \( E[Z_1Z_2] := \rho \), we obtain

\[
E(Z_1Z_2)^2 = 1 + 2\rho^2.
\]

Now \( E(T_iT_2) = E \prod_{i=1}^{2} \xi^{-1}(\alpha_iZ_i) \) and since \( ET_i = \beta_i \left( 1 + \frac{\alpha_i^2}{2} \right) \), \( Var(T_i) = (\alpha_i\beta_i)^2 \left( 1 + \frac{5}{4}\alpha_i^2 \right) \) hence

\[
Cov(T_1, T_2) \geq \beta_1\beta_2 \left( \frac{\alpha_1\alpha_2\rho}{2} \right)^2 \quad \text{and} \quad Corr(T_1, T_2) \geq \frac{\alpha_1\alpha_2\rho^2}{\sqrt{1 + \frac{5}{4}\alpha_1^2 \sqrt{1 + \frac{5}{4}\alpha_2^2}}}.\]

The exact expression leads to complex integrals the determination of which requires numerical integration.

**A MULTIVARIATE DISTRIBUTION OF COMPONENT LIVES**

**A PRÉCIS OF COHERENT STRUCTURES**

The concept of coherent structures has been introduced by Birnbaum et.al. (1961) according to which a coherent structure function, say \( \phi \), is a monotone increasing binary function of some given number, say \( n \), of binary variates such that we have

\[
\phi(0,\ldots,0) = 0, \quad \phi(1,\ldots,1) = 1
\]

and in which there are no irrelevant components.

Here the binary variates are indicators of component operation while the structure function, \( \phi \), indicates under any specific subset of component operations, whether the system operates or not.
Some common examples of such functions are:

\[ \phi(x_1, x_2, \cdots, x_n) = \prod_{i=1}^{n} x_i \quad \text{a series system} \]

\[ \phi(x_1, x_2, \cdots, x_n) = 1 - \prod_{i=1}^{n} (1 - x_i) \quad \text{a parallel system} \]

\[ \phi(x_1, x_2, \cdots, x_n) = \{ \sum_{i=1}^{n} x_i \geq k \} \quad \text{a k-out-of-n system} \]

Here we introduce notation for the indicator function for any event \( A \) as \( \{ A \} \).

If \( T_i \) is the life-length of the \( j^{th} \) component then the indicator variate for it being alive at any time \( t \) is

\[ \{ T_i > t \} = \begin{cases} 1 & \text{if } T_i > t, \\ 0 & \text{if } T_i \leq t. \end{cases} \]

Thus the reliability of the system at time \( t > 0 \) is given by the expectation of the structure, viz.,

\[ E[\{ T_1 > t \}, \cdots, \{ T_n > t \}] = R(t), \]

where

\[ E\{ T_i > t \} = \Pr[T_i > t] = F_i(t) := 1 - F_i(t). \]

NB: \( R(t) \neq \phi[\bar{F}_1(t), \cdots, \bar{F}_n(t)] \) because in the algebraic expansion of \( \phi(x) \) there may be squared terms \( x_i^2 \) but since all the \( x_i \) are idempotent (i.e. \( x_i^2 = x_i \)), we cannot have \( \bar{F}_i^2 = \bar{F}_i \). So simple substitution fails.

Let \( \psi_i \) be the \( j^{th} \) min-path of the structure \( \phi \) of which there are a number \( p \) of components. (Recall that a path of a structure is a subset of its components which if all are operational will make the system operational; a min-path is a path such that if any one of its components fails the system will fail.) A complex system may have 30,000 components to be considered, so there are as many as \( 2^{30,000} \) subsets and the determination of all min-paths is often a very difficult task.

Let \( \psi_j(x) = \prod_{i \in \psi_j} x_i \) be the series subsystem representing the \( j^{th} \) min-path. Then the structure function can be expressed as

\[ \phi(x) = \prod_{j=1}^{p} \psi_j(x) = 1 - \prod_{j=1}^{p} [1 - \psi_j(x)] \]

Thus every structure can be represented as a parallel system of series min-paths.

**THE CASE OF MULTI-SYSTEM FAILURE**

However the state of a component cannot always be well-described as either "on" or "off", especially in biological life where malfunctioning organs (components) results in "sickness" not "death". Similar arguments can be made for many other complex systems. This would be especially true when multiple system components have lives determined by cumulative damage (fatigue).
Consider a life length
\[ T > t = \sum_{i=1}^{N} X_i < \omega \approx [Z < \frac{1}{\alpha_i} \xi(t/\beta_i)] \]

We do not need a binary indicator structure function but one that allows the system to malfunction but not fail, when the operation of some of its components is degraded but they have not yet failed. This should be a function of each component’s cumulative damage:
\[ \Phi(z_1, \ldots, z_n) \quad \text{where} \quad z_i = \frac{1}{\alpha_i} \xi(t_i/\beta_i) \]

In many situations, it may be appropriate to extend the definition of the Fatigue Life distribution to the case of multi-system failure through \( \Phi \), the standard multivariate normal (Gaussian) cdf. The multi-variate Gaussian (Normal) distribution is the unique multi-variate extension of the (univariate) Gaussian family of variates which is the only family of random variables which is closed under arbitrary linear combinations. That is, if \( X \) and \( Y \) are real Gaussian random variables then \( aX + bY \) is also a univariate real Gaussian variate for all real \( a, b \). A real Gaussian variate is the result of a sum of a sufficiently large number of small random increments, either positive or negative, that are uniformly negligible. The Gaussian variates may also be characterized as the class of variates which are closed under convolution and have a virtually linear increasing hazard rate. Here it is the hyperbolic transformations of the scaled time, at which the cumulative damage from the general strain repetitions at each site cause the damage-tolerance to be almost simultaneously exceeded, which are assumed to be multivariate Gaussian. This assumption would not apply in all situations.

**Definition:** If the \( P = (\rho_{i,j}) \) is a correlation matrix with \( P^{-1} = (s_{i,j}) \) and there exist vectors \( \alpha = (\alpha_1, \ldots, \alpha_n) \) and \( \beta = (\beta_1, \ldots, \beta_n) \) such that for the function \( \xi \), as defined previously, the random vector of \( \xi - \text{transformations} \) of the failure times has the specific distribution:
\[ \left( \frac{1}{\alpha_1} \xi(T_1/\beta_1), \ldots, \frac{1}{\alpha_n} \xi(T_n/\beta_n) \right) \sim \mathcal{N}(0, \Sigma) \]

Now to evaluate it, one must determine (or estimate) not only the scale and shape parameters for each component life but also its correlation with all other components as well. Then one must likely employ computer programs, and perhaps, Mathematica for the integration.

**THE PLATEAU EFFECT IN THE MULTI-SITE MODEL**

We now consider a site where all \( m \) components have the same life and share equally a part of the global stress. The probability that all will survive until time \( t \) may be approximated by
\[ [\bar{F}(t/m)]^m = e^{m \ln \bar{F}(t/m)} = e^{-mH(t/m)} = e^{-t/mH(t/m)} \]

Since \( H \) is the hazard function, hence \( H' = h \) is the hazard rate. We note that if we let \( \frac{t}{m} = u \), we have
Habibullah and Saunders

\[ \frac{H(u)}{u} = \frac{1}{u} \int_0^u h(t) \, dt \]

which is the average hazard rate and

\[ \left[ \tilde{F}(t/m) \right]^m = e^{-\frac{H(u)}{u}}. \]

It is the asymptotic behavior of the average hazard rate which is important. When \( \alpha < 1 \), the fatigue lives are IHRA, and, as such, are well-adapted to model the lives of machines as well as biological organisms. (IHRA, rather than just IHR is the wide class of life distributions of great interest in Reliability. IHR is the natural class of lives of all coherent structures.)

We examine the asymptotic behavior of the univariate Fatigue Life hazard rate. For \( T \sim F\mathcal{L}(\alpha, \beta) \) then the hazard function and the hazard rate are

\[ H(t) = -\ln \Phi_{\alpha}^{1/\alpha} \left[ \xi(t/\beta) \right], \quad h(t) = \frac{\varphi_{\alpha}^{1/\alpha} \left[ \xi(t/\beta) \right]}{\Phi_{\alpha}^{1/\alpha} \left[ \xi(t/\beta) \right]} \times \psi(t/\beta) \]

Recall Mills’ classical ratio:

\[ M(x) := \frac{x \varphi(x)}{\varphi(x)} = 1 - \frac{1}{x^2} + \frac{3}{x^4} - \cdots \quad \text{for } x \gg 1.\]

After some algebraic simplification, we have

\[ h(t) = \left[ \frac{1 - (\beta/t)^2}{2\alpha^2 \beta} \right] \times \frac{1}{M[\alpha^{1/\alpha} \xi(t/\beta)]}. \]

We note

\[ \frac{1}{\alpha} \xi(t/\beta) \gg 1 \quad \text{iff} \quad \frac{t}{\beta} \gg \left[ \frac{\alpha}{2} + \sqrt{1 + \left( \frac{\alpha}{2} \right)^2} \right]^2 \]

and

\[ \left[ \tilde{F}(t/m) \right]^m \asymp e^{\psi \left( \frac{t}{2\alpha^2 \beta} \right)} \quad \text{for } t \gg \beta. \]

Hence, we see that under many asymptotic conditions the multi-site model will have a life that exhibits the late-life plateau, namely, that the hazard rate becomes asymptotically constant.

**A FEW REMARKS ON THE SIGNIFICANCE OF THE MVBS DISTRIBUTION**

We begin with an account of some recent papers relating to the bivariate and multivariate Birnbaum Saunders distribution.

**The MVBS: A New Domain**

Diaz-Garcia and Dominguez-Molina (2006) present univariate and multivariate generalizations of the two-parameter BS distribution, the three-parameter BS distribution and the sinh-normal distributions; present some alternative definitions under the assumption of stochastic representation of a spherical random vector, including the log-elliptical distribution; emphasize that all distributions derived belong to the family of spherical distributions.
The section entitled "Applications" shows some graphs of densities of the univariate generalized three parameter Birnbaum-Saunders distribution, the log-elliptical distribution and the sinh-spherical distribution; utilizes data on the fatigue life of 6061-T6 aluminum coupons given in birnsaun69b to illustrate the way in which some of the univariate distributions derived in the preceding sections can be estimated.


Kundu et al. (2010) introduce the bivariate Birnbaum Saunders distribution whose marginals are univariate BS distributions; an absolutely continuous bivariate distribution and can have either positive or negative correlation; five unknown parameters; various properties of BVBS distribution; non-linear equations for obtaining maximum likelihood estimators; likelihood ratio tests for some hypotheses of interest; asymptotic distributions of the maximum likelihood estimators; confidence intervals for parameters using the asymptotic distribution of the MLE; Monte Carlo simulations to examine the performance of the proposed estimators.

Numerical data analysis to illustrate the estimation methods proposed; Paired data on the bone mineral density (BMD) measured in $g = cm^2$ for 24 individuals, who had participated in an experimental study. (Observations: (1.103 1.027), (0.842 0.857), (0.925 0.875), $\cdots$) The first figure represents the BMD of the bone Dominant Radius before starting the study and the second figure represents the BMD of the same bone after one year. Use of the BVBS distribution to model these bivariate data; computation of the Kolmogorov Smirnov (KS) distances between the empirical marginals and the fitted marginals.

The authors conclude that the BVBS distribution is indeed a good model for the BMD bivariate data; compute the 95 percent probability coverages of confidence intervals based on the pivotal quantities associated with the estimators obtained from the empirical Fisher information matrix; comment that the asymptotic confidence intervals do not work well when the sample size is very small, as the coverage probabilities are much lower than the nominal level, but, the performance is quite satisfactory for large sample sizes; comment that the multivariate Birnbaum-Saunders distribution also can be discussed along the same lines, and that several properties of the BVBS distribution can be extended to the multivariate case as well, but the inference becomes complicated; comment that work is currently under progress in this direction.

Jørgensen’s paper under review (submitted in 2011) entitled "Construction of Multivariate Dispersion Models" is available on the net; in this paper, the author emphasizes on the need to develop flexible multivariate distribution families for stochastic modelling of non-normal data; the author remarks: “There is a large variety of such families available, including multivariate hyperbolic distributions (1987), skew-normal and skew-elliptical distributions (2010), and multivariate Birnbaum-Saunders distributions (2006; 2010), to name but a few......It is not easy to know where to turn, as the following comment by Letac (2007) illustrates: `While the names of distributions in
The Potential of the MVBS Distribution

From the preceding discussion it is clear that the MVBS is not another in the "jungle" of multivariate distributions. The derivation of the MVBS being based on the physical realities of components and systems that undergo cumulative damage, the distribution is able to model the multivariate plateau effect. However, the worth of a newly derived model being demonstrable only when an exact practical situation is modeled, the applicability of the MVBS to related fatigue-lives is yet to be explored. (A fairly detailed application of the BVBS distribution to paired data on bone mineral density (BMD) published in the 2010 paper by Kundu et al. (2010) is the only empirical application available to date.) The MVBS distribution seems to carry great potential for modeling the fatigue-lives of related components, especially, multi-site fatique, in an industrial setting.

ACKNOWLEDGEMENT

• The research work of Dr. Saleha Naghmi Habibullah was supported by the Higher Education Commission, Pakistan through Sabbatical Leave Fellowship Award No. 20-8/SLFP/HEC/R&D/2011 by which she was Visiting Scholar at the University of Washington from July 18 to Oct 17, 2011.

• The authors wish to thank Chris Glazner for his advice and assistance in the use of LaTeX in the presentation of the graphs computed and drawn using Mathematica.

REFERENCES

29. Weibull, V. (1967). *Personal Communication to Sam C. Saunders*
ECONOMIC DESIGN OF TIME TRUNCATED SAMPLING PLANS FOR WEIBULL DISTRIBUTION USING PERCENTILES AS QUALITY PARAMETER

Muhammad Aslam\textsuperscript{1}, Muhammad Azam\textsuperscript{1}, Chi-Hyuck Jun\textsuperscript{2} and Munir Ahmad\textsuperscript{3}

\textsuperscript{1} Department of Statistics, Forman Christian College University, Lahore, Pakistan. Email: aslam_ravian@hotmail.com
\textsuperscript{2} Department of Industrial and Management Engineering, POSTECH, Pohang 790-784, Republic of Korea. Email: chjun@postech.ac.kr
\textsuperscript{3} National College of Business Administration and Economics Lahore, Pakistan. Email: drmunir@ncbae.edu.pk

ABSTRACT

In this paper, acceptance sampling plans based on truncated life tests are proposed when the lifetime of the product follows the Weibull distribution. The proposed plan is developed using the percentiles life and a cost model is established. The optimal values of plan parameters are determined so as to minimize the total cost of the experiment while both producer’s risk and consumer’s risk are satisfied. The true fraction defective is modeled using the Bayesian approach. Extensive tables are provided for practical use. The examples are given to illustrate the plan developed here.

KEY WORDS

Cost model; Bayesian approach; acceptance sampling plans; life tests; producer and consumer risks.

1. INTRODUCTION

Acceptance sampling is used for acceptance or rejection of the lots of products through inspection of the product. The decision through the inspection of the lot is based on two numbers namely; sample size and acceptance number. For example, the product is rejected when the number of failures/non-conforming items is larger than the acceptance number. So, the choice of these two numbers is very important at the time of inspection. Especially, in life testing experiments, the cost is directly attached with the sample size. Cost is large for the big sample and low for small sample. These two parameters are determined using the acceptance sampling schemes such as attribute sampling, variable sampling, repetitive group sampling and resubmitted sampling schemes. In this modern age, latest computer software for example Analyzer (2007) and Software program TP105 (2008) are available to find these plan parameters by satisfying producer’s risk (the probability of rejecting a good product) and consumer’s risk (the probability of accepting a bad lot) but as pointed out by Hsu (2009) that all the computer provides the sampling plans without cost consideration. Further, the producer and consumer wants the maximum protection while using the acceptance sampling schemes as well as the low experiment cost of life tests. The acceptance sampling based on time truncated life tests have been studied by many authors including Baklizi (2004), Rosaiah et al. (2006), Tsai and Wu (2006), Balakrishnan et al. (2007), Lio et al. (2010 a, b) and Aslam et al. (2010)
Economic Design of Time Truncated Sampling Plans


According to the best of the author knowledge, there is no study on attribute sampling using the cost model for the time truncated life tests using the Bayesian approach. In this paper, we will consider the attribute acceptance sampling for time truncated life tests using the Bayesian approach when the lifetime of the product follows the Weibull distribution with known or unknown shape parameters. We will use the cost model to find the plan parameters using the percentiles life of the Weibull distribution as quality parameter. The rest of the paper is organized as economic design along with the cost model is given in Section 2. Analysis and discussion is given in Section 3. In last section some concluding remarks are given.

2. ECONOMIC DESIGN OF PROPOSED PLAN

Let \( t_q \) be the true percentile life of the product, \( t_{q0} \) be the true percentile life of the product, \( \alpha \) is the producer’s risk and \( \beta \) is the consumer’s risk. The lot is accepted if the true percentile life is greater than the specified percentile life. Otherwise the lot is rejected. The sampling plan is proposed as

**Step-1** Select a random sample of size \( n \) form the lot and put on the test for specified experiment time \( t \). Decide about the acceptance number \( c \).

**Step-2** Accept the lot if the number of failure are smaller than or equal to \( c \) during the experiment time, but reject the lot, otherwise.

The proposed plan is characterized by two plan parameters such as \( n \) and \( c \). Suppose that the lifetime of the product under inspection follows the Weibull distribution with cumulative distribution function (cdf) given as

\[
F(t) = 1 - \exp \left( - \left( \frac{t}{\theta} \right)^{\omega} \right), \quad t \geq 0
\]

where \( \omega \) is the shape parameter and \( \theta \) is the scale parameter of the Weibull distribution. The q-th percentile of the Weibull distribution is given by

\[
y_q = \theta \left( \ln \left( \frac{1}{1-q} \right) \right)^{1/\omega}
\]

The OC function of the plan is given as

\[
L(p) = \sum_{i=0}^{c} \binom{n}{i} p^i (1 - p)^{n-i}
\]

where \( p \) is the function of the cdf of the Weibull distribution. In acceptance sampling plans based on truncated life tests, it is convenient to write the experiment time equal to
Aslam, Azam, Jun and Ahmad

termination ratio $\alpha$ and specified percentile life $t_{q_0}$ i.e., $t_0 = at_{q_0}$. Based on this, the cdf given in Eq. (1) can be written as

$$p = 1 - \exp \left[ -\alpha \left( \frac{t_q}{t_{q_0}} \right) \right]$$

(4)

2.1 Proposed Plan with Cost Constraint

A plan in which items are classified into two categories i.e. good or bad based on the information of random sample selected from the lot is called the attribute acceptance sampling plan. According to plan stated above the lot of the product is accepted if the number of recorded failures are less than the specified failures $c$ from $n$ items, otherwise the lot is rejected. Hence, the proposed plan test the simple hypothesis $H_0 = t_q \geq t_{q_0}$ versus $H_1 = t_q < t_{q_0}$. Let $\alpha$ be the producer’s risk, $\beta$ be the consumer’s risk. We need to find the plan parameters of the proposed plan such that

$$L(p_1) \geq 1 - \alpha$$

(5)

and

$$L(p_2) < \beta$$

(6)

where $p_1$ is the probability of failure of an item before experiment time $t_0$ at the producer’s risk usually at specified percentiles ratio $t_q/t_{q_0}$ and $p_2$ be the probability of failure of an item before experiment time $t_0$ at consumer’s risk usually at ratio $t_q/t_{q_0} = 1$. So, to find the plan parameters of the plan $n$ and $c$, one needs to use the Eq. (5) and Eq. (6). The values of $n$ and $c$ affect the cost time truncated life test. In fact, in life testing experiment, the budget of such experiment is limited. Therefore, the designing of the proposed attribute acceptance sampling plans under cost constraint is very important issue to the experimenter, industrialist and reliability analyst. Let $TC(n,c)$ denote the total cost the experiment that is consists of following three parts [for more detail reader may refer to Hsu, 2009].

$C_i$: Inspection cost per item

$C_f$: Internal failure cost

$C_0$: The cost of an outgoing defective unit, i.e., the post-sale failure.

Therefore, the total cost of the time truncated life test is

$$TC(n,c) = C_i ATI + C_f D_d + C_0 D_n$$

(7)

where $ATI$ is average total inspection, $D_d$ denote the defective items detected, $AOQ$ is average outgoing quality and $D_n$ denote the defective items not detected during the life test, these are given as

$$AOQ = \frac{pL(p)(N-n)}{N}$$

(8)

where $AOQ$ is average outgoing quality.

$$ATI = n + \left( 1 - L(p) \right)(N - n)$$

(9)

$$D_d = np + \left( 1 - L(p) \right)(N - n)p$$

(10)

$$D_n = L(p)(N - n)p$$

(11)
2.2 Bayesian Approach

Hsu (2009) proposed the attribute single acceptance sampling plan by fixing the value of \( p \) in Eq. (8) to Eq. (11). But in practice, usually, \( p \) given in above equations is unknown in advance. So, to find the values of \( p \) in these equations is questionable. Bayesian approach is used to handle this type of problem in practice. It is well known theory that the natural conjugate of the Weibull distribution is beta distribution. Therefore, the usual prior of \( p \) is beta distribution. So, to find the cost of the plan we need to integrate the function of \( AOQ, ATI, D_d \) and \( D_n \) using the probability density function (pdf) of the beta distribution. To simplify the equations in Bayesian approach, we will use numerical method of integration. Let \( w(p) \) is the pdf of the beta distribution given by

\[
w(p) = \frac{p^{\delta-1}(1-p)^{\tau-1}}{B(\delta, \tau)}
\]

where \( \delta > 0 \) and \( \tau > 0 \) are shape parameters of the beta distribution. The functions of \( AOQ, ATI, D_d \) and \( D_n \) can be written as follows

\[
AOQ = \int_0^1 \frac{p L(p)(N-n)p^{\delta-1}(1-p)^{\tau-1}}{B(\delta, \tau)} dp
\]

(13)

\[
ATI = \int_0^1 [n + (1 - L(p))(N - n)] \frac{p^{\delta-1}(1-p)^{\tau-1}}{B(\delta, \tau)} dp
\]

(14)

\[
D_d = \int_0^1 [np + (1 - L(p))(N - np)] \frac{p^{\delta-1}(1-p)^{\tau-1}}{B(\delta, \tau)} dp
\]

(15)

\[
D_n = \int_0^1 [L(p)(N - np)] \frac{p^{\delta-1}(1-p)^{\tau-1}}{B(\delta, \tau)} dp
\]

(16)

We need to integrate the above functions using the numerical methods of evaluating. After integrating the above functions by method of numerical evaluating the simplified form of these functions are given as

\[
AOQ = \frac{N-n}{N} \sum_{i=0}^n \binom{n}{i} \times \frac{\Gamma(\delta+i)(n+\tau-i-1)(\delta+\tau-1)!}{\Gamma(\delta-1)(\tau-1)(n+\delta+\tau)!}
\]

\[
ATI = n + (N - n) \sum_{i=c+1}^n \binom{n}{i} \times \frac{\Gamma(\delta+i)(n+\tau-i-1)(\delta+\tau-1)!}{\Gamma(\delta-1)(\tau-1)(n+\delta+\tau)!}
\]

\[
D_d = \frac{n \text{Beta}(\delta+1, \beta)}{\text{Beta}(\delta, \beta)} + (N - n) \sum_{i=c+1}^n \binom{n}{i} \times \frac{\Gamma(\delta+i)(n+\tau-i-1)(\delta+\tau-1)!}{\Gamma(\delta-1)(\tau-1)(n+\delta+\tau)!}
\]

\[
D_n = (N - n) \sum_{i=0}^n \binom{n}{i} \times \frac{\Gamma(\delta+i)(n+\tau-i-1)(\delta+\tau-1)!}{\Gamma(\delta-1)(\tau-1)(n+\delta+\tau)!}
\]

We assume the shape parameters \( \tau = \delta = 2 \) for illustration the procedure developed here. Any other values of shape parameters can be in above derived functions (17) to (20). For given values of shape parameters above functions can be written as follows

\[
AOQ = \frac{N-n}{N} \sum_{i=0}^n \binom{n}{i} \times \frac{(2+\tau)(n+\tau-i)(3)!}{(2-1)(2-1)(n+4)!}
\]

(21)

\[
ATI = n + (N - n) \sum_{i=a+1}^n \binom{n}{i} \times \frac{(i+1)(n+1-i)(3)!}{(2-1)(2-1)(n+3)!}
\]

(22)

\[
D_d = \frac{n \text{Beta}(3, 2)}{\text{Beta}(2, 2)} + (N - n) \sum_{i=a+1}^n \binom{n}{i} \times \frac{(2+\tau)(n+\tau-i)(3)!}{(2-1)(2-1)(n+4)!}
\]

(23)

\[
D_n = (N - n) \sum_{i=0}^n \binom{n}{i} \times \frac{(2+\tau)(n+\tau-i)(3)!}{(2-1)(2-1)(n+4)!}
\]

(24)
The economic sampling plan can be found through the following mathematical model

Minimize $TC(n, c) = C_i \cdot ATI + C_f \cdot D_d + C_0 \cdot D_n$ (25a)

Subject to

\[ \sum_{i=0}^{c} \binom{n}{i} p_1 (1 - p_1)^{n-i} \geq 1 - \alpha \] (25b)

\[ \sum_{i=0}^{c} \binom{n}{i} p_2 (1 - p_2)^{n-i} \leq \beta \] (25c)

So, we will find the plan parameters by minimizing the $TC(n, c)$ for $\delta = \tau = 2$. The proposed plan is characterized by two parameters namely acceptance number and sample size. We will find these parameters such that Eq. (25a) to Eq. (25c) should satisfy. The posterior distribution is considered as the Weibull distribution. Suppose that the lifetime of the product follows the Weibull distribution with known or unknown shape parameter. The cumulative distribution function (cdf) of this distribution is given as

\[ F(t) = 1 - \exp\left(-\left(\frac{t}{\lambda}\right)^\gamma\right), t \geq 0 \] (26)

where $\gamma$ is the shape parameter and $\lambda$ is the scale parameter. If the shape parameters are unknown it can be estimated from the previous failure time data. It is important to note that the cdf of the Weibull distribution depends on $\lambda$ only through $t/\lambda$.

The $qth$ percentile of the Weibull distribution is given by

\[ \gamma_q = \theta \left( \ln \left( \frac{1}{1-q} \right) \right)^{1/\omega} \] (27)

It is more convenient to express the experiment time $t_0$ as the multiple of specified percentile life $t_{q0}$ and termination ratio $\alpha$. So, the probability of failure $p$ in Eq. (26) through Eq. (27) are given by

\[ p = 1 - \exp \left[ -\alpha^\omega \left( \frac{t_q}{t_{q0}} \right)^\omega \ln \left( \frac{1}{1-q} \right) \right] \] (28)

2.3 Algorithm

We used R program to find the plan parameters of the proposed plan using the following algorithm

**Input:** $a=0.5$, $\gamma =2$ (shape parameters), $p=0.5$ (50th percentile), $alpha=2$, $Beta=2$, $C_i=1$, $C_f=2$, $C_0=4$

- Computation of probabilities $p1$ and $p2$ using Weibull distribution
- $p1=1-\exp((-1)^*a1^b*(ratio)^\gamma*(-b)^log((1/(1-p)), \exp(1)))$
- $p2=1-\exp((-1)^*a1^b*log((1/(1-p)), \exp((1))) \quad \text{# at ratio=1}$
- $AOQ = c(); Dd = c(); Dn = c(); ATI = c(); TC = c(); TCmin = c()$
- for (i in 1:J) {
  - $a=sample(0:10, nos, replace=TRUE); n=sample (1:100, nos, replace=TRUE) \quad \text{# where “nos” is the number of samples (W.R) at each iteration.}$
  - $wa = which(a < n); l = length(wa)$
  - $LP1 = pbinom(a, n, p1, TRUE); LP2 = pbinom(a, n, p2, TRUE)$
  - $AOQ <- NULL; Dn <- NULL; Dd <- NULL; ATI <- NULL; TC <- NULL$
Economic Design of Time Truncated Sampling Plans

```r
for (k in 1:l) {
  Computation of AOQ[k], Dd[k], Dn[k], ATI[k] and TC[k] as mentioned above
}
w1 = which(LP1 >= 0.95); w2 = which(LP2 <= β); m = match(w1,w2)
Then corresponding AOQ, Dd, Dn, ATI, LP1 and LP2
}
TCmin[i] = min(TC) && nmin[i] = min(n)
Then corresponding AOQ, Dd, Dn, ATI, LP1 and LP2
}
Selection of minimum value of TC from all iterations & corresponding variables
AOQ, Dd, Dn, ATI, n, a, LP1 and LP2
}
final_output = cbind(ATI, Dd, Dn, AOQ, n, a, TCmin, Lp1, LP2)
Output: return (final_output)
```

3. ANALYSIS AND DISCUSSION

To illustrate the proposed plan we will conduct some numerical study. We set the various values of shape parameter of the Weibull distribution ($γ = 1, 2, 3$), various percentile values (10%, 20% and 50%), various values of termination ratio ($α = 0.5, 1.0$), producer’s risk is 5%, different values of consumer’s risk ($β = 0.25, 0.10, 0.05, 0.01$) and various values of percentiles ratio ($t_q / t_{q0} = 2, 4, 6, 8$). We presented the plan parameters, AOQ, ATI, $D_d$, $D_n$ and total minimum cost in these tables also for $γ = 1, 2$ and 3 in Tables 1-3 when percentile value is 50% and $α = 0.5$.

**Table 1: Plan Parameters when $w=1$, $α=1.0$ and 20% percentiles**

<table>
<thead>
<tr>
<th>$β$</th>
<th>$t_q / t_{q0}$</th>
<th>$α=1.0$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ATI</td>
<td>Dd</td>
</tr>
<tr>
<td>0.25</td>
<td>2</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>928.57</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>928.65</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>928.50</td>
</tr>
<tr>
<td>0.10</td>
<td>2</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>950.53</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>954.15</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>955.95</td>
</tr>
<tr>
<td>0.05</td>
<td>2</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>958.33</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>964.91</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>968.35</td>
</tr>
<tr>
<td>0.01</td>
<td>2</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>968.28</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>973.09</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>977.63</td>
</tr>
</tbody>
</table>
```

Note: (-) shows parameter does not exist.
Table 2: Plan Parameters when \( w=2, \alpha=1.0 \) and 20% percentiles

<table>
<thead>
<tr>
<th>( \beta )</th>
<th>( t_q/t_q^0 )</th>
<th>( a=1.0 )</th>
<th>ATI</th>
<th>Dd</th>
<th>Dn</th>
<th>AOQ</th>
<th>n</th>
<th>c</th>
<th>TC</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.25</td>
<td>2</td>
<td>928.57</td>
<td>490.25</td>
<td>9.75</td>
<td>0.00975</td>
<td>25</td>
<td>3</td>
<td>1948.07</td>
<td></td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>929.50</td>
<td>489.01</td>
<td>10.99</td>
<td>0.01099</td>
<td>13</td>
<td>1</td>
<td>1951.48</td>
<td></td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>933.80</td>
<td>487.96</td>
<td>12.04</td>
<td>0.01204</td>
<td>7</td>
<td>0</td>
<td>1957.87</td>
<td></td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>933.80</td>
<td>487.96</td>
<td>12.04</td>
<td>0.01204</td>
<td>7</td>
<td>0</td>
<td>1957.87</td>
<td></td>
</tr>
<tr>
<td>0.10</td>
<td>2</td>
<td>950.53</td>
<td>494.67</td>
<td>5.33</td>
<td>0.00533</td>
<td>45</td>
<td>5</td>
<td>1961.19</td>
<td></td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>959.39</td>
<td>495.10</td>
<td>4.90</td>
<td>0.00490</td>
<td>18</td>
<td>1</td>
<td>1969.19</td>
<td></td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>967.40</td>
<td>495.65</td>
<td>4.35</td>
<td>0.00435</td>
<td>11</td>
<td>0</td>
<td>1976.09</td>
<td></td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>983.78</td>
<td>498.84</td>
<td>1.16</td>
<td>0.00116</td>
<td>31</td>
<td>1</td>
<td>1987.09</td>
<td></td>
</tr>
<tr>
<td>0.05</td>
<td>2</td>
<td>955.95</td>
<td>495.94</td>
<td>4.06</td>
<td>0.00406</td>
<td>57</td>
<td>6</td>
<td>1966.45</td>
<td></td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>971.51</td>
<td>497.09</td>
<td>2.91</td>
<td>0.00291</td>
<td>22</td>
<td>1</td>
<td>1977.33</td>
<td></td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>971.51</td>
<td>497.09</td>
<td>2.91</td>
<td>0.00291</td>
<td>22</td>
<td>1</td>
<td>1977.33</td>
<td></td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>978.25</td>
<td>497.58</td>
<td>2.42</td>
<td>0.00242</td>
<td>14</td>
<td>0</td>
<td>1983.08</td>
<td></td>
</tr>
<tr>
<td>0.01</td>
<td>2</td>
<td>968.28</td>
<td>497.35</td>
<td>2.65</td>
<td>0.00265</td>
<td>83</td>
<td>8</td>
<td>1973.58</td>
<td></td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>980.58</td>
<td>498.50</td>
<td>1.50</td>
<td>0.00150</td>
<td>39</td>
<td>2</td>
<td>1983.58</td>
<td></td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>984.78</td>
<td>498.84</td>
<td>1.16</td>
<td>0.00116</td>
<td>31</td>
<td>1</td>
<td>1987.09</td>
<td></td>
</tr>
</tbody>
</table>

Table 3: Plan Parameters when \( w=3, \alpha=1.0 \) and 20% percentiles

<table>
<thead>
<tr>
<th>( \beta )</th>
<th>( t_q/t_q^0 )</th>
<th>( a=1.0 )</th>
<th>ATI</th>
<th>Dd</th>
<th>Dn</th>
<th>AOQ</th>
<th>n</th>
<th>c</th>
<th>TC</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.25</td>
<td>2</td>
<td>929.50</td>
<td>489.01</td>
<td>10.99</td>
<td>0.01099</td>
<td>13</td>
<td>1</td>
<td>1951.48</td>
<td></td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>933.80</td>
<td>487.96</td>
<td>12.04</td>
<td>0.01204</td>
<td>7</td>
<td>0</td>
<td>1957.87</td>
<td></td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>933.80</td>
<td>487.96</td>
<td>12.04</td>
<td>0.01204</td>
<td>7</td>
<td>0</td>
<td>1957.87</td>
<td></td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>933.80</td>
<td>487.96</td>
<td>12.04</td>
<td>0.01204</td>
<td>7</td>
<td>0</td>
<td>1957.87</td>
<td></td>
</tr>
<tr>
<td>0.10</td>
<td>2</td>
<td>967.40</td>
<td>495.65</td>
<td>4.35</td>
<td>0.00435</td>
<td>11</td>
<td>0</td>
<td>1976.09</td>
<td></td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>967.40</td>
<td>495.65</td>
<td>4.35</td>
<td>0.00435</td>
<td>11</td>
<td>0</td>
<td>1976.09</td>
<td></td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>967.40</td>
<td>495.65</td>
<td>4.35</td>
<td>0.00435</td>
<td>11</td>
<td>0</td>
<td>1976.09</td>
<td></td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>967.40</td>
<td>495.65</td>
<td>4.35</td>
<td>0.00435</td>
<td>11</td>
<td>0</td>
<td>1976.09</td>
<td></td>
</tr>
<tr>
<td>0.05</td>
<td>2</td>
<td>978.25</td>
<td>497.58</td>
<td>2.42</td>
<td>0.00242</td>
<td>14</td>
<td>0</td>
<td>1983.08</td>
<td></td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>983.78</td>
<td>498.84</td>
<td>1.16</td>
<td>0.00116</td>
<td>31</td>
<td>1</td>
<td>1987.09</td>
<td></td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>983.78</td>
<td>499.15</td>
<td>0.8513</td>
<td>0.00085</td>
<td>21</td>
<td>0</td>
<td>1991.06</td>
<td></td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>983.78</td>
<td>499.15</td>
<td>0.8513</td>
<td>0.00085</td>
<td>21</td>
<td>0</td>
<td>1991.06</td>
<td></td>
</tr>
</tbody>
</table>

From Tables 1-3, we note following trends in plan parameters.
1. When percentiles ratio increases from 2 to 8, the sample size decreases.
2. The acceptance number is decreased as the percentiles ratio increase from 2 to 8.
3. Total minimum cost is increased as the percentiles ratio increases.
4. In Table 1, for percentile ratio 1, we could not find the values of plan parameters to satisfy the conditions.
5. There is no any specific trend in other plan parameters.
6. When the shape parameter increases from 1 to 3, the total cost is increases. For example, when \( \gamma = 1 \) and \( t_q/t_{q_0} = 4 \), the total minimum cost is 1948.07, for \( \gamma = 2 \), it is 1951.48 and for \( \gamma = 3 \), it is 1957.87.

Tables 4-5 are presented for two values of shape parameter (\( \gamma = 2, 3 \)), \( \alpha = 0.1 \) and percentile value 10%. When compare the total minimum cost of the experiment for percentile 10% to 20%, we note that for all other same specified parameters when the percentile value decreases from 20 to 10% the minimum total cost of the experiment increase. For example, for all other same specified values and percentile ratio 4, as the percentile point changes from 20% to 10%, the total minimum cost increases from 1951.49 to 1983.71.

### Table 4: Plan Parameters when \( w = 2, \alpha = 1.0 \) and 10% percentiles

<table>
<thead>
<tr>
<th>( \beta )</th>
<th>( t_q/t_{q_0} )</th>
<th>( \alpha = 1.0 )</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ATI</td>
<td>Dd</td>
</tr>
<tr>
<td>0.25</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>980.87</td>
<td>498.62</td>
</tr>
<tr>
<td>4</td>
<td>980.35</td>
<td>498.31</td>
</tr>
<tr>
<td>6</td>
<td>978.25</td>
<td>497.58</td>
</tr>
<tr>
<td>8</td>
<td>978.25</td>
<td>497.58</td>
</tr>
<tr>
<td>0.10</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>987.37</td>
<td>499.29</td>
</tr>
<tr>
<td>4</td>
<td>989.62</td>
<td>499.34</td>
</tr>
<tr>
<td>6</td>
<td>989.62</td>
<td>499.34</td>
</tr>
<tr>
<td>8</td>
<td>990.22</td>
<td>499.25</td>
</tr>
<tr>
<td>0.05</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>4</td>
<td>992.80</td>
<td>499.62</td>
</tr>
<tr>
<td>6</td>
<td>992.80</td>
<td>499.62</td>
</tr>
<tr>
<td>8</td>
<td>994.13</td>
<td>499.64</td>
</tr>
<tr>
<td>0.01</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>4</td>
<td>995.33</td>
<td>499.82</td>
</tr>
<tr>
<td>6</td>
<td>996.23</td>
<td>499.85</td>
</tr>
<tr>
<td>8</td>
<td>996.23</td>
<td>499.85</td>
</tr>
</tbody>
</table>

Note: (-) shows parameter does not exist.
Table 5: Plan Parameters when \( w=3, \alpha=1.0 \) and 10\% percentiles

<table>
<thead>
<tr>
<th>( \beta )</th>
<th>( t_q / t_{q0} )</th>
<th>ATI</th>
<th>Dd</th>
<th>Dn</th>
<th>AOQ</th>
<th>n</th>
<th>c</th>
<th>TC</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.25</td>
<td>2</td>
<td>980.35</td>
<td>498.31</td>
<td>1.6853</td>
<td>0.00169</td>
<td>27</td>
<td>1</td>
<td>1983.72</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>978.25</td>
<td>497.58</td>
<td>2.4167</td>
<td>0.00242</td>
<td>14</td>
<td>0</td>
<td>1983.08</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>978.25</td>
<td>497.58</td>
<td>2.4167</td>
<td>0.00242</td>
<td>14</td>
<td>0</td>
<td>1983.08</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>978.25</td>
<td>497.58</td>
<td>2.4167</td>
<td>0.00242</td>
<td>14</td>
<td>0</td>
<td>1983.08</td>
</tr>
<tr>
<td>0.10</td>
<td>2</td>
<td>988.80</td>
<td>499.34</td>
<td>0.6646</td>
<td>0.00066</td>
<td>52</td>
<td>2</td>
<td>1990.13</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>990.22</td>
<td>499.25</td>
<td>0.7523</td>
<td>0.00075</td>
<td>22</td>
<td>0</td>
<td>1991.73</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>990.22</td>
<td>499.25</td>
<td>0.7523</td>
<td>0.00075</td>
<td>22</td>
<td>0</td>
<td>1991.73</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>990.22</td>
<td>499.25</td>
<td>0.7523</td>
<td>0.00075</td>
<td>22</td>
<td>0</td>
<td>1991.73</td>
</tr>
<tr>
<td>0.05</td>
<td>2</td>
<td>991.80</td>
<td>499.58</td>
<td>0.4195</td>
<td>0.00042</td>
<td>61</td>
<td>2</td>
<td>1992.64</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>994.13</td>
<td>499.64</td>
<td>0.3559</td>
<td>0.00036</td>
<td>29</td>
<td>0</td>
<td>1994.84</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>994.13</td>
<td>499.64</td>
<td>0.3559</td>
<td>0.00036</td>
<td>29</td>
<td>0</td>
<td>1994.84</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>994.13</td>
<td>499.64</td>
<td>0.3559</td>
<td>0.00036</td>
<td>29</td>
<td>0</td>
<td>1994.84</td>
</tr>
<tr>
<td>0.01</td>
<td>2</td>
<td>994.64</td>
<td>499.79</td>
<td>0.2118</td>
<td>0.00021</td>
<td>97</td>
<td>3</td>
<td>1995.06</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>996.23</td>
<td>499.85</td>
<td>0.1477</td>
<td>0.00015</td>
<td>64</td>
<td>1</td>
<td>1996.52</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>997.35</td>
<td>499.89</td>
<td>0.1105</td>
<td>0.00011</td>
<td>44</td>
<td>0</td>
<td>1997.57</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>997.35</td>
<td>499.89</td>
<td>0.1105</td>
<td>0.00011</td>
<td>44</td>
<td>0</td>
<td>1997.57</td>
</tr>
</tbody>
</table>

Note: (-) shows parameter does not exist.

We also consider the median life (50\% percentiles) of the product to find the plan parameters. The plan parameters for median are found for various values of termination ratio \( \alpha = 0.5, 1.0 \), various values of shape parameters and placed in Tables 6-11. We note the same trends in plan parameters as in Tables 1-5.

Table 6: Plan Parameters when \( w=1, \alpha=0.5 \) and 50\% percentiles

<table>
<thead>
<tr>
<th>( \beta )</th>
<th>( t_q / t_{q0} )</th>
<th>ATI</th>
<th>Dd</th>
<th>Dn</th>
<th>AOQ</th>
<th>n</th>
<th>c</th>
<th>TC</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.25</td>
<td>2</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>862.04</td>
<td>474.13</td>
<td>25.87</td>
<td>0.02587</td>
<td>17</td>
<td>3</td>
<td>1913.77</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>866.05</td>
<td>474.08</td>
<td>25.92</td>
<td>0.02592</td>
<td>13</td>
<td>2</td>
<td>1917.89</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>866.05</td>
<td>474.08</td>
<td>25.92</td>
<td>0.02592</td>
<td>13</td>
<td>2</td>
<td>1917.89</td>
</tr>
<tr>
<td>0.10</td>
<td>2</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>896.71</td>
<td>484.04</td>
<td>15.9589</td>
<td>0.01596</td>
<td>30</td>
<td>5</td>
<td>1928.62</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>903.26</td>
<td>484.72</td>
<td>15.2848</td>
<td>0.01528</td>
<td>21</td>
<td>3</td>
<td>1933.83</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>913.77</td>
<td>486.45</td>
<td>13.5501</td>
<td>0.01355</td>
<td>17</td>
<td>2</td>
<td>1940.87</td>
</tr>
<tr>
<td>0.05</td>
<td>2</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>911.56</td>
<td>487.55</td>
<td>12.4537</td>
<td>0.01245</td>
<td>38</td>
<td>6</td>
<td>1936.47</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>923.25</td>
<td>489.16</td>
<td>10.8444</td>
<td>0.01084</td>
<td>24</td>
<td>3</td>
<td>1944.94</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>923.25</td>
<td>489.16</td>
<td>10.8444</td>
<td>0.01084</td>
<td>24</td>
<td>3</td>
<td>1944.94</td>
</tr>
<tr>
<td>0.01</td>
<td>2</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>930.17</td>
<td>491.44</td>
<td>8.5649</td>
<td>0.00856</td>
<td>55</td>
<td>8</td>
<td>1947.30</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>945.68</td>
<td>493.72</td>
<td>6.2762</td>
<td>0.00628</td>
<td>36</td>
<td>4</td>
<td>1958.23</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>951.42</td>
<td>494.49</td>
<td>5.5057</td>
<td>0.00551</td>
<td>31</td>
<td>3</td>
<td>1962.43</td>
</tr>
</tbody>
</table>

Note: (-) shows parameter does not exist.
**Table 7:** Plan Parameters when \( w=1, \alpha=1.0 \) and 50% percentiles

<table>
<thead>
<tr>
<th>( \beta )</th>
<th>( t_q/t_{q0} )</th>
<th>( a=1.0 )</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ATI</td>
<td>Dd</td>
</tr>
<tr>
<td>0.25</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>-</td>
</tr>
<tr>
<td>4</td>
<td>663.43</td>
<td>405.00</td>
</tr>
<tr>
<td>6</td>
<td>669.40</td>
<td>402.20</td>
</tr>
<tr>
<td>8</td>
<td>669.40</td>
<td>402.20</td>
</tr>
<tr>
<td>0.10</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>-</td>
</tr>
<tr>
<td>4</td>
<td>734.42</td>
<td>434.47</td>
</tr>
<tr>
<td>6</td>
<td>761.14</td>
<td>441.64</td>
</tr>
<tr>
<td>8</td>
<td>765.76</td>
<td>441.09</td>
</tr>
<tr>
<td>0.05</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>-</td>
</tr>
<tr>
<td>4</td>
<td>756.22</td>
<td>443.12</td>
</tr>
<tr>
<td>6</td>
<td>788.50</td>
<td>451.27</td>
</tr>
<tr>
<td>8</td>
<td>788.50</td>
<td>451.27</td>
</tr>
<tr>
<td>0.01</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>-</td>
</tr>
<tr>
<td>4</td>
<td>794.66</td>
<td>456.86</td>
</tr>
<tr>
<td>6</td>
<td>834.38</td>
<td>467.04</td>
</tr>
<tr>
<td>8</td>
<td>862.04</td>
<td>474.13</td>
</tr>
</tbody>
</table>

Note: (-) shows parameter does not exist.

**Table 8:** Plan Parameters when \( w=2, \alpha=0.5 \) and 50% percentiles

<table>
<thead>
<tr>
<th>( \beta )</th>
<th>( t_q/t_{q0} )</th>
<th>( a=0.5 )</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ATI</td>
<td>Dd</td>
</tr>
<tr>
<td>0.25</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>954.15</td>
<td>494.95</td>
</tr>
<tr>
<td>4</td>
<td>955.16</td>
<td>494.33</td>
</tr>
<tr>
<td>6</td>
<td>945.89</td>
<td>490.98</td>
</tr>
<tr>
<td>8</td>
<td>945.89</td>
<td>490.98</td>
</tr>
<tr>
<td>0.10</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>968.36</td>
<td>497.26</td>
</tr>
<tr>
<td>4</td>
<td>973.70</td>
<td>497.41</td>
</tr>
<tr>
<td>6</td>
<td>973.70</td>
<td>497.41</td>
</tr>
<tr>
<td>8</td>
<td>978.25</td>
<td>497.58</td>
</tr>
<tr>
<td>0.05</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>973.45</td>
<td>497.92</td>
</tr>
<tr>
<td>4</td>
<td>981.62</td>
<td>498.47</td>
</tr>
<tr>
<td>6</td>
<td>981.62</td>
<td>498.47</td>
</tr>
<tr>
<td>8</td>
<td>985.97</td>
<td>498.72</td>
</tr>
<tr>
<td>0.01</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>987.92</td>
<td>499.26</td>
</tr>
<tr>
<td>4</td>
<td>990.12</td>
<td>499.39</td>
</tr>
<tr>
<td>6</td>
<td>990.12</td>
<td>499.39</td>
</tr>
</tbody>
</table>

Note: (-) shows parameter does not exist.
Table 9: Plan Parameters when $w=2$, $a=1.0$ and 50% percentiles

<table>
<thead>
<tr>
<th>$\beta$</th>
<th>$t_q/t_{q0}$</th>
<th>$a=1.0$</th>
<th>ATI</th>
<th>Dd</th>
<th>Dn</th>
<th>AOQ</th>
<th>n</th>
<th>c</th>
<th>TC</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.25</td>
<td>2</td>
<td>663.43</td>
<td>405.00</td>
<td>95.0000</td>
<td>0.09500</td>
<td>12</td>
<td>4</td>
<td>1853.43</td>
<td></td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>715.71</td>
<td>417.08</td>
<td>82.9167</td>
<td>0.08292</td>
<td>5</td>
<td>1</td>
<td>1881.55</td>
<td></td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>700.60</td>
<td>400.20</td>
<td>99.8000</td>
<td>0.09980</td>
<td>2</td>
<td>0</td>
<td>1900.20</td>
<td></td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>700.60</td>
<td>400.20</td>
<td>99.8000</td>
<td>0.09980</td>
<td>2</td>
<td>0</td>
<td>1900.20</td>
<td></td>
</tr>
<tr>
<td>0.10</td>
<td>2</td>
<td>734.42</td>
<td>434.47</td>
<td>65.5333</td>
<td>0.06553</td>
<td>17</td>
<td>5</td>
<td>1865.48</td>
<td></td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>817.95</td>
<td>456.37</td>
<td>43.6318</td>
<td>0.04363</td>
<td>7</td>
<td>1</td>
<td>1905.21</td>
<td></td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>817.95</td>
<td>456.37</td>
<td>43.6318</td>
<td>0.04363</td>
<td>7</td>
<td>1</td>
<td>1905.21</td>
<td></td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>857.71</td>
<td>464.43</td>
<td>35.5714</td>
<td>0.03557</td>
<td>4</td>
<td>0</td>
<td>1928.86</td>
<td></td>
</tr>
<tr>
<td>0.05</td>
<td>2</td>
<td>756.22</td>
<td>443.12</td>
<td>56.8826</td>
<td>0.05688</td>
<td>21</td>
<td>6</td>
<td>1869.98</td>
<td></td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>849.70</td>
<td>466.93</td>
<td>33.0667</td>
<td>0.03307</td>
<td>8</td>
<td>1</td>
<td>1915.83</td>
<td></td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>849.70</td>
<td>466.93</td>
<td>33.0667</td>
<td>0.03307</td>
<td>8</td>
<td>1</td>
<td>1915.83</td>
<td></td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>849.70</td>
<td>466.93</td>
<td>33.0667</td>
<td>0.03307</td>
<td>8</td>
<td>1</td>
<td>1915.83</td>
<td></td>
</tr>
<tr>
<td>0.01</td>
<td>2</td>
<td>794.66</td>
<td>456.86</td>
<td>43.1392</td>
<td>0.04314</td>
<td>30</td>
<td>8</td>
<td>1880.94</td>
<td></td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>881.10</td>
<td>478.25</td>
<td>21.7500</td>
<td>0.02175</td>
<td>14</td>
<td>2</td>
<td>1924.60</td>
<td></td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>907.62</td>
<td>483.70</td>
<td>16.3022</td>
<td>0.01630</td>
<td>11</td>
<td>1</td>
<td>1940.23</td>
<td></td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>907.62</td>
<td>483.70</td>
<td>16.3022</td>
<td>0.01630</td>
<td>11</td>
<td>1</td>
<td>1940.23</td>
<td></td>
</tr>
</tbody>
</table>

Note: (-) shows parameter does not exist.

Table 10: Plan Parameters when $w=3$, $a=0.5$ and 50% percentiles

<table>
<thead>
<tr>
<th>$\beta$</th>
<th>$t_q/t_{q0}$</th>
<th>$a=0.5$</th>
<th>ATI</th>
<th>Dd</th>
<th>Dn</th>
<th>AOQ</th>
<th>n</th>
<th>c</th>
<th>TC</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.25</td>
<td>2</td>
<td>985.65</td>
<td>498.94</td>
<td>1.0599</td>
<td>0.00106</td>
<td>32</td>
<td>1</td>
<td>1987.77</td>
<td></td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>982.74</td>
<td>498.27</td>
<td>1.7263</td>
<td>0.00173</td>
<td>16</td>
<td>0</td>
<td>1986.19</td>
<td></td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>982.74</td>
<td>498.27</td>
<td>1.7263</td>
<td>0.00173</td>
<td>16</td>
<td>0</td>
<td>1986.19</td>
<td></td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>982.74</td>
<td>498.27</td>
<td>1.7263</td>
<td>0.00173</td>
<td>16</td>
<td>0</td>
<td>1986.19</td>
<td></td>
</tr>
<tr>
<td>0.10</td>
<td>2</td>
<td>992.30</td>
<td>499.62</td>
<td>0.3820</td>
<td>0.00038</td>
<td>63</td>
<td>2</td>
<td>1993.07</td>
<td></td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>993.29</td>
<td>499.57</td>
<td>0.4329</td>
<td>0.00043</td>
<td>27</td>
<td>0</td>
<td>1994.16</td>
<td></td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>993.29</td>
<td>499.57</td>
<td>0.4329</td>
<td>0.00043</td>
<td>27</td>
<td>0</td>
<td>1994.16</td>
<td></td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>993.29</td>
<td>499.57</td>
<td>0.4329</td>
<td>0.00043</td>
<td>27</td>
<td>0</td>
<td>1994.16</td>
<td></td>
</tr>
<tr>
<td>0.05</td>
<td>2</td>
<td>994.40</td>
<td>499.76</td>
<td>0.2386</td>
<td>0.00024</td>
<td>74</td>
<td>2</td>
<td>1994.88</td>
<td></td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>995.88</td>
<td>499.79</td>
<td>0.2112</td>
<td>0.00021</td>
<td>35</td>
<td>0</td>
<td>1996.30</td>
<td></td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>995.88</td>
<td>499.79</td>
<td>0.2112</td>
<td>0.00021</td>
<td>35</td>
<td>0</td>
<td>1996.30</td>
<td></td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>995.88</td>
<td>499.79</td>
<td>0.2112</td>
<td>0.00021</td>
<td>35</td>
<td>0</td>
<td>1996.30</td>
<td></td>
</tr>
<tr>
<td>0.01</td>
<td>2</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>997.46</td>
<td>499.92</td>
<td>0.0825</td>
<td>0.00008</td>
<td>78</td>
<td>1</td>
<td>1997.63</td>
<td></td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>998.22</td>
<td>499.94</td>
<td>0.0613</td>
<td>0.00006</td>
<td>54</td>
<td>0</td>
<td>1998.34</td>
<td></td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>998.22</td>
<td>499.94</td>
<td>0.0613</td>
<td>0.00006</td>
<td>54</td>
<td>0</td>
<td>1998.34</td>
<td></td>
</tr>
</tbody>
</table>

Note: (-) shows parameter does not exist.
Table 11: Plan Parameters when \( w=3 \), \( \alpha=1.0 \) and 50\% percentiles

<table>
<thead>
<tr>
<th>( \beta )</th>
<th>( t_q/t_{q_0} )</th>
<th>( \alpha=1.0 )</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ATI</td>
<td>Dd</td>
</tr>
<tr>
<td>0.25</td>
<td>2</td>
<td>669.00</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>700.60</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>700.60</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>700.60</td>
</tr>
<tr>
<td>0.10</td>
<td>2</td>
<td>765.76</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>857.71</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>857.71</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>857.71</td>
</tr>
<tr>
<td>0.05</td>
<td>2</td>
<td>788.50</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>849.70</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>893.39</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>893.39</td>
</tr>
<tr>
<td>0.01</td>
<td>2</td>
<td>862.04</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>907.62</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>933.80</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>933.80</td>
</tr>
</tbody>
</table>

Note: (-) shows parameter does not exist.

3.1 Example
Suppose that a quality engineer in manufacturer wants to adopt the proposed plan for lot of product having \( N=1000 \) units under inspection. Suppose that the lifetime of the product follows the Weibull distribution with shape parameter 2. Let \( \alpha = 0.05, \beta = 0.05, t_q/t_{q_0} = 4, C_i=1, C_f=2 \) and \( C_0=4 \).

He wants to test the product for 500 hours. Let the median life of the product is 1000 hours. From Table, we have

\[ ATI=981.62, \quad D_d=498.47, \quad D_n=1.5272, \quad AOQ=0.00153, \quad n=28, \quad c=1 \] and total minimum cost is 1984.67. The plan is implemented as put 28 items on the test for 500 hours. If the number of failures is larger than 2, then reject the product otherwise accept the product. The total minimum cost of the experiment is 1984.67.

4. COMMENTS AND CONCLUSION
The economic acceptance sampling plan is considered in the paper using the Bayesian approach based on time truncated life tests. Extensive tables are given for industrial use. The plan parameters are found by minimizing the cost model. The tables given in the paper can be used for the testing purpose of any electronic product such as the automobiles, components of airplanes, charger of mobiles, ball bearing products etc. The present approach can be extended to develop the group acceptance sampling plans using the cost model. This paper is under process. The present approach can also be extended using the double sampling, sequential sampling and two stage group sampling schemes as a future research. Some other distributions can be used to find the plan parameters as a future research.
REFERENCES

ROLE OF LINE MANAGER IN IMPLEMENTING HUMAN RESOURCE MANAGEMENT AND HUMAN RESOURCE DEVELOPMENT

Madeeha Jabeen
Department of Computer Sciences
The Islamia University of Bahawalpur, Pakistan
Email: madeeha141@hotmail.com

ABSTRACT

This study reveals the role of line manager in implementing practices regarding people management and development. To date line manager role is devolved to line HR (human resource) responsibilities. Due to changes in organizational structure and culture and also increased awareness and need for developmental activities in organizations line manager is contributing a lot in learning and development of employees. HR- Line partnership along with the benefits to both are discussed. Author has tried to develop a conceptual framework elaborating effect of devolved line manager role on areas such as learning and development, organizational development, decision making, understanding of individual and organizational need and effective resolution of workplace problems by line manager. All these factors are being positively affected by line manger involvement in HR activities which in turn enhances organizational effectiveness and performance.

KEYWORDS

Human Resource Management; Human Resource Development; Line manager; devolved line manager role; HR-line relationship.

1. INTRODUCTION

The practices involved in from hiring to firing and retiring of employees comes under the umbrella of human resource management. The planning/formulation of strategies is more in hand of HR managers and the execution of these strategies are made viable through line managers. Human resource management (HRM) and Human resource development (HRD) go hand in hand in organization in order to enhance performance. Development of employees address the areas of orientation, coaching, mentoring, training and learning, performance management, career development, organizational development and change and employs counseling of employees include. (Werner, 2006).

Role of line manager HR responsibilities is area of research from a decade. Before that in recent year’s line manager’s role to implement HRM and HRD is not apparent. Staff manager is mostly concerned with the development practices for human resource. But currently this trend has been changed, as devolution of responsibilities from staff manger to line manager is taking place. Therefore functions which are strategic and operational in nature are responsibility of line managers. The light has been shed by many researches in
past years on importance of line managers’ role within HR practices (Brown, 2008). In future the importance and influence of relationship between HR and line managers will be maintained (Faisal, 2011).

1.1 Human resource management

Human resource management is “The policies and practices involved in carrying out ‘people’ or human resource aspect of management position including recruiting, screening, training, appraising and rewarding.” (Dessler, 2000).

In contrast According to Werner (2006), primary HRM functions are human resource planning, equal employment opportunity, staffing, compensation and benefits, employee (labor) relations, health, safety and security, and human resource development. Whereas secondary HRM functions are organization/job design, performance management and performance appraisal systems, research and information systems. Which can also categorized in five basic practices of HRM that are recruitment & selection, training & development, performance management, compensation management, employee relations.

1.2 Human resource development activities

Ongoing Progression of employees through a series of stages, with different set of issues, themes, and duties is known as career development. Which is further split into two areas: career planning and career management. Coaching, mentoring, training and learning, performance management, career development, organizational development and change and employs counseling are important developmental framework for HRD (Werner, 2006).

2. DEVOLUTION OF HR ACTIVITIES TO LINE MANAGER

Because of the shift in focus from personnel problems to strategic HR issues, it is at this point where HRM finds itself very close to the territory of HRD. At present HRM functions have shifted within organizations from Personnel to line-management, therefore HRD functions have often followed go well with HRD being, which in earlier times considered simply as a sub-function of HRM (Wright, 2000).

According to McCarthy (2010), Opportunities and challenges are being faced due to delegation of decisions regarding human resource made by line management. More responsibility and power for managing staff is exercised by line manager due to active involvement in HR related issues. On the other hand, significant challenges are faced by them and researches signify that contradictions in implementation of HR policies are act out by line managers. In contrast Suff (2011), suggests that most of the people management practices on Day-to-day basis is now decentralized to line managers.

3. ROLE OF LINE MANAGER IN HUMAN RESOURCE DEVELOPMENT

“A line manager is defined as a manager who is authoritative to direct the work of subordinates and is conscientious for achieving organizational goals”. Line functions plays role in directly accomplishing organizational goals Dessler, (2000). In contrast “Line authority is given to mangers and organizations that are directly responsible for the production of goods and services” Line authority supplant staff authority matters related
to production of goods and services (Werner, 2006). The primary belief for line manager is that by recognizing, developing and sustaining the suitable and required competencies and commitment, they achieve the organization’s strategic objectives and performance through driving HRD policies and practice.

Line managers encourages work teams to share knowledge among them which is becoming source of capturing tacit knowledge and as transformational leaders they influence experience, attitude and personalities which ultimately motivates them to share knowledge among teams and work groups. Another work done by line manager as intermediary is that line managers stipulate work between senior management and employees and act as channel for transmitting information between them and to execute the decisions taken by senior managers They are responsible to persuade decisions made by top management without direct authority (McNeil, 2003).

In contrast Garavan (1991), McCracken (2000), &Harrison (2002), states learning on the job is becoming very important with increase in competition, so now both working and learning go hand in hand. Informal learning is also essential learning activity. Now the view of learning is changing from traditional HRD activities to learning on the job which are being executed by line managers to make workplace enriched with learning. Due to strategic nature of developmental activities HRD experts keep an eye on the work standards along with providing guidance, developmental activities are executed by line managers because of devolution of HRD activities to operational managers.

The work relationship between line manger and HRD connoisseur arises from the change in nature of work for HRD specialist. HRD connoisseurs are concerned with strategic issues in spite of day to day and work as change agent by implementing strategic and operational change. As they should be authorized as HRD advisor rather HRD providers, HRD specialist must provide assistance to line manager when and where it is required, and helping them in problems regarding learning plans, particular skills and competencies required to make them viable. All these circumstances stimulates line manger to work for HRD functions. (McCracken, 2000; Hockey, 2005; & Garavan, 1991).

Harrison (2002), emphasizes that to enhance learning and development line manger should be part of planning, and to delegate key responsibility to the line. Significant changes in organizations, groups and individual level as well as in human resource management (HRM) functions, more association with learning and development of line staff has been intended that embers most present-day interest.

Whereas in contras Hay (2002), for example, argues that people are not satisfied with development of talent and skills, regardless of their satisfaction with pay which is becoming one of the most vigilant factor due to which talented people are leaving from organizations. The term HRM must be used in a neutral sense; to mean the whole of people management, the combination of employee resourcing, employee relations, employee reward and learning &development issues in business and management that together comprises the field of HRM. Learning & development is the activity performed by both line and HR manager rather than isolating it to HR specialist, as its functions starts from hiring to firing and retiring of employees and recruitment, rewarding,
developing relations etc comes under the umbrella of HRM therefore the term HRM is used for both line and staff functions.

Line manager can be a good coach? The answer to this question is elaborated by Matthew, (2011) that good coaching and training can be given by internal manager? There are not ample results provided by the professional coaches. Yes this can be done by them which results in exchange of dialogues to stimulate culture which provides assistance, support and mutual trust. To date it is need for organizations to promote and encourage coaching culture which will definitely affect the performance. So if line manager will be supporting and encouraging people who are in direct contact with him, significant improvement will be realized among employees.

Owing to bad Liaoning with the line managers results in dissatisfaction regarding skills and competencies development, the people with extraordinary performance are departing from organizations. Under this scenario training and development need and activities are being performed by line managers for smoothing the progress of employees learning and development (Hay, 2000). It is not the responsibility of line manager to play dual role of assessor and coach as it is also not formally included in line managers’ performance objectives. More familiarity of needs related organization, individuals as well as business context by line manager than HRD specialist, are capable of dealing with these learning and development imperative needs. As they are more vigilant to day to day operations and customers, this is an importance way which explores knowledge gap and develops understanding about major organizational issues and realities and by line managers which is not apparent to HR manager.

To increase organizational performance through HRD it is essential to link systematic training needs with performance appraisal process. Need for training arises from long term goals, technology advancement, and nature of work. So at operational level linking needs with performance appraisal process is critical for which line manager’s responsible. due to strategic swing of HRD functions, expectations from line manager role in HRD is also changing which increases their responsibility and accountability for human resource development. In order to make sure to carry new roles by HRD specialist and line manager large amount of investment is needed in capacity development of both.(Šiugždinien, 2008)

Mc Mullen (2008), sheds light on line managers role in implementing performance management as HR task. Line manager also plays significant role in performance management and had considerable effect on it. Performance management process requires central role by line managers. Line manager must know how to translate organizational goals into departmental goals by cascading them from upper to lower level. It is the line manager as being the ultimate supervisor can understand needs of employees what they have to perform for organizational success as stated earlier, employees must answerable to line manager for what they are doing and manager must clarify role designs to employees so that they can held accountable for their work. The performance management objectives are fulfilled if this process goes on continuous basis along with constructive feedback taken by line manager in order to know where they actually stand and what interventions can be made to overcome gaps between actual and desired performance (Mc Mullen, 2008).
In contrast Bos-Neles (2010), argues that the perceived relevance of the tasks and responsibilities, along with the quantity and quality of the tasks determine the desire for line manager to perform HR tasks. If they are empathetic about tasks and responsibility related to HR the feeling that their HR tasks and responsibilities are relevant, and realize that it can bring added value to the performance of their team, can be helpful for employees in motivating them or and for their wellbeing, then they are more enthusiastic to perform HR tasks and will certainly put more effort into their people management role.

4. INCREASE LINE MANAGERS INVOLVEMENT IN HRM AND HRD FUNCTIONS; BENEFITS TO ORGANIZATION

In organizations a number of benefits are visualized due to devolved HR responsibilities to line managers. Benefit for HR experts is more freedom to connect with strategic issues which ultimately add value to organization. Secondly good relationship between line and HR managers is developed, creating partnership which helps in managing employees. Another benefit to line manager is that they can easily deal with complex issues related to HR due to their increased involvement for which they are admired for and encouraged. So they are more indulged in workplace management decisions taking place on daily basis by taking voluntary actions which in result speedily and proactively resolve issues and problems arising day to day. These are some of the enablers of line manager HR involvement.

Whereas opposite to these are some inhibitors of line manager HR involvement given by McGuire (2008), according to him it can be problematic to held line manager responsible for HR issues why? Because the major issue is lack of requires skills or insufficient training regarding HR activities like day to day performance management processes, bargaining processes, and compensation management by line manager. Priorities of line manager also changes due to devolution of HR responsibilities which may decrease value of HR. Resulting in threats to HR standards throughout organization. Conflicts are resolved more quickly and employees are retained due to devolvement of HR responsibility to line managers (David McGuire, 2008).

In contrast Gibb (2003), states line manager attitude towards HR is also changing along with human relations at workplace and organizational change creating Liaoning between them due to more involvement in HR issues. Organizational performance and effectiveness is increased due to line manger participation in communication, guidance and coaching. (Hutchinson and Purcell, 2003).There is need for training at workplace rather than in formal education to promote lifelong learning of employees. Line manager are in direct contact with employees daily operations so they capable of understanding learning needs more effectively. Line staff involvement in development and erudition of employees results in several benefits. Key role is played by line manager to increase learning of employees by forecasting their needs regarding development. Continuous support by line manager is provided to enhance learning and development. Another benefit is that it helps in diminishing gap between organizational and individual needs because line manager is more familiar with individual needs at workplace rather HR specialist who may be unaware of both organizational and individual needs. Very obvious benefit to line manager due to involvement in learning and development and is
their own learning and people management skills. Other competencies developed are interpersonal skills which help in managing teams and group and become proficient in handling problems regarding people. Due to these competencies positive impact among organization is seen enhancing overall management in organizations.

The final advantage frequently identified is organizational development by contribution to organizational change and revolutionize human relations at work because of line managers role as develop of people. They have aligned human relations with innovativeness, quality and efficiency generating value for money. Through learning and development partnership between line and HR managers is established resulting in Liaoning between them and creating environment of mutual trust and team work. (Gibbs, 2003).

In contrast Jamie, propose that due to hyper competition and organizational change development of HRM to line manager is increasing partnership between HR Specialist and line manager. This partnership is impacting both positively and increasing organizational effectiveness and performance. However Gibbs (2003), also point out disadvantages of line manager role as developer that is criticism of more participation by line management in helping employees to learn at work. In short involvement of line manager can enhance performance in organizations creating mutual fulfillments for managers, learners and organizations. Through learning environment which fosters knowledge sharing.

5. DEVELOPING THE FRAMEWORK

This study explores the line manager’s role in implementing/executing activities related to HRM and HRD. As HRM and developmental practices becomes burning need for today’s hypercompetitive environment therefore people involved in formulation and execution of strategies and practices significantly affects performance by playing vital roles. In order to implement HRM and HRD practices line manager’s role is increasing day by day and is more vigilant as compare to past. As line manger mediate, confer and interprets connections between strategic and operational levels of organization, further light should be shed on his role in implementing HR practices.
Figure 1 is a framework showing relationship between role of line manager in HRM & HRD and its impact on different activities related to HR, which is ultimately affecting performance in positive way. This model aims at exploring vital role played by line manager in executing human development practices, which is being devolved from staff manager. Win-Win approach is seen between line manager and HR specialist creating benefits for each other. This relationship is called as partnership. Benefit to line manager is that he has the opportunity to polish his skills and competencies with the help of HR Specialist. Clearly understanding organizational and individual needs by line manager, organizational development, quick decision making, increased learning and development, effectively resolving problems at workplace are some of the areas which have been positively affected due to devolution of line manager role in HR functions, which ultimately enhances performance of organization.

Devolution of HR responsibilities to line manager involve learning and developmental tasks, understanding and executing training needs, Partnership between HR specialist and line managers which in result create good relationship among them and increased learning of line managers as well, speedier decision making, more quicker in resolving day to day issues due to direct interactions of line manager with employees.
yielding increase in performance. Line manager due to its close interaction with employees can understand organizational and individual needs more wisely as compare to HR specialist. As they are more concerned with their day to day activities so they can easily understand their developmental needs. In this way their role in implementing HRD practices is vigilant nowadays. Learning and development is enhanced by line manager which ultimately increases performance of both. Due to vital role played by line manager in learning and development their own grooming of skills is also taking place which is benefiting both employees and line manager.

6. CONCLUSION

This study broadens the understanding of role of line manager in implementing HRM & HRD. Devolved HR role of line manager is drastically increasing and is playing a vital role in learning, developing and coaching of employees. HR line Partnership is win-win relation giving benefits to both parties resulting in organizational effectiveness. Some areas such as employee increase learning and development, increased decision making, organizational development, effective resolution of workplace problems are areas positively influenced by devolved HR role. Performance of organizations increases due to devolved line manager HR responsibilities. Different researches are consistent with this study as Hay (2000), shed light on role of line manager as developer. Benefits to organization due to devolved line manager role of HR activities are discussed by McGuire (2008), and Gibbs (2003).

7. LIMITATION

To investigate empirically the devolved role of line manager in HRM and HRD is needed although data on actual implementation of the idea is insufficient.

8. FUTURE IMPLICATIONS AND SUGGESTIONS

Better use of technology can be useful to support and smooth people management process by line managers. Just only wishing that line authority will be more accountable and HR People more strategic is not sufficient, the need of time is to implement wise solutions regarding HR-Line collaborative work and roles should be divided in order to achieve goals. (Brown, 2008).This study can be beneficial for organizations with devolved HR role of line manager and its effect on performance. Helps them in understanding how they can create value in their organizations through Win-Win approach of line and HR manager. Relationship of line manager with HR specialist must be enhanced and also they must strengthen this relation by providing ongoing support, assistance and guidance to each other. This can fill the gap of knowledge and basic developmental needs between them. This will be having significant affect on employees as well as organizations performance. Future researches can be done by empirically testing the model proposed in this study along with the factors which affects line manager’s role in implementing HRM and HRD.
REFERENCES

ESTIMATION OF THE SCALE AND LOCATION PARAMETERS BASED UPON THE UPPER RECORDS VALUES OF WEIBULL DISTRIBUTION FROM COMPLETE DATA, TYPE II SINGLY AND DOUBLY CENSORED DATA

Riffat Gabeen¹, Azaz Ahmad² and Abdul Samad Hirai³
¹ College of Statistics and Actuarial Science, Punjab University, Lahore, Pakistan. Email: riffat.jabeen79@gmail.com
² Govt. Postgraduate College, Okara, Pakistan. Email: azazpu@yahoo.com
³ 52 Mamdot Block, Mustafa Town, Lahore, Pakistan

SUMMARY

In this paper we discuss the Estimation of location and scale parameters based on the Upper record values of Weibull distribution by the best linear unbiased estimator (BLUE) and an alternative linear estimates (Gupta) from complete samples as well as from Type II singly and doubly censored samples. Tables of coefficients for best linear estimator and alternative estimator of \( \mu \) and \( \lambda \) are presented for various choices of censoring for \( n \leq 8 \). Variances and covariances of estimator of \( \mu \) and \( \lambda \) for BLUE and variances of the estimator of \( \mu \) and \( \lambda \) for Alternative estimator are also presented. The computational formula and procedure used are explained.

1. INTRODUCTION

Records are not only memorable in their times but also reflect the progress in science and technology and forecast the evaluation of mankind on the basis of record achievements in different areas of activity for example people are very excited about records regarding the geography and the various natural or social phenomena. This importance of the records motivates the necessity to construct mathematical models of records and to develop the corresponding mathematical theory and to fit the models to real process and also forecast the scale of the next record earthquake or floods and to suggest possible measures to take adequate precautions. 1st paper on mathematical theory of record that appeared was by chandler in 1952. After this lot of work has been done in this field, interested reader may refer to Glick (1978), Galambos (1978), Resnick (1987), Nagarja (1988), Nevzorov (1987), Ahsanullah (1988), Arnold and Balakrishan (1989). Ahsanullah, M (2004) has discussed the estimation of the location and scale parameters based upon upper record values of Rayleigh distribution for complete sample without giving the Tables of coefficients of the estimates and variance covariances. Also Ahsanullah, M (2004) has discussed the estimation of the location and scale parameters based upon upper record values of Weibull distribution for complete sample without giving the Tables of coefficients of the estimates and variance covariances.

2. UPPER RECORD VALUES FROM WEIBULL DISTRIBUTION

In large group of problems where the occurrence of an event in any part of object may be said to have occurred in the object as a whole for example the phenomena of yield
limits. Structural or dynamic strength, electrical insulation, break downs, life of electrical bulbs or even the death of a man as the probability of surviving depends on the probability of non died from many different causes. We have to satisfy the function $\varphi(x)$ and the necessary general condition this function has to satisfy is to be one non-decreasing function. The most general and simple function satisfying the condition is

$$\left(\frac{x-a}{b}\right)^c$$ and thus we put

$$F(x) = 1 - \exp\left(\frac{x-a}{b}\right)^c \quad x > 0, a > 0, b > 0, c > 0$$

which is Weibull cumulative distribution function (c.d.f).

The c.d.f of weibull distribution of wide applicability fits the observations better than other known distribution functions in life testing context. It is used as a tolerance limit in the analysis of quantum response data. It is also used for the data of cooking utensils which have long lives.

A record is an entry that is smaller or greater than all previous entries. If an entry that is greater then all previous entries then we call it upper record value.

Suppose $x_1, x_2, x_3, ..., x_n$ be a sequence of independent and identical distributed random variable from the Weibull distribution whose reduced density function is

$$f(x) = x^{\gamma-1} \exp\left(-\frac{x^\gamma}{\gamma}\right) \quad 0 < x < \infty$$

$$F(x) = 1 - \exp\left(-\frac{x^\gamma}{\gamma}\right)$$

then $y_n = \max\{x_1, x_2, x_3, ..., x_n\}$ for $n \geq 1$ we call $x_j$ is the upper record value if $y_j > y_{j-1}, j > 1$.

Let us denote the $E(y_{U(n)}) = \alpha_n, \text{Cov}(y_{U(n)}, y_{U(m)}) = \omega_{n,m} - \alpha_n \alpha_m$ and $\text{Var}(y_{U(n)}) = \omega_{n,n} (\alpha_n)^2$ has been tabulated in Table I and Table II for sample size $n$ up to order 8.

Further let $y_{U(r+1)}, y_{U(r+2)}, y_{U(r+3)}, ..., y_{U(n-r-r)}$ is the sequence of upper record values from Weibull distribution, whose density function is

$$f(x) = \frac{1}{\Gamma r} \frac{x^{3\gamma-1} e^{-x^{\gamma}/3}}{3^{\gamma-1}} \quad x > 0$$

We take $\gamma = 3$ (assuming shape parameter as known) are available for Type II censoring. In which $r_1$ and $r_2$ are fixed in advance. $r_1$ is the missing observation on left and $r_2$ is the missing observations on right. $(n-r_1-r_2)$ are measured observation. So
$r_1$ is the smallest and $r_2$ is the largest observations were not observed due to experimental restrictions. Under such condition we have to estimates these parameters by Lloyd’s and Gupta method.

### Table I

<table>
<thead>
<tr>
<th>n</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>Means</td>
<td>1.287900</td>
<td>1.717200</td>
<td>2.0034</td>
<td>2.225999</td>
<td>2.411501</td>
<td>2.572264</td>
<td>2.715175</td>
<td>2.844462</td>
</tr>
</tbody>
</table>

### Table II

<table>
<thead>
<tr>
<th>s/r</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.219099</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>0.135657</td>
<td>0.180877</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>0.102371</td>
<td>0.136495</td>
<td>0.159245</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>0.083942</td>
<td>0.111923</td>
<td>0.130576</td>
<td>0.145085</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>0.072016</td>
<td>0.096021</td>
<td>0.112025</td>
<td>0.124472</td>
<td>0.134845</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>0.063588</td>
<td>0.084785</td>
<td>0.098915</td>
<td>0.109906</td>
<td>0.119065</td>
<td>0.127002</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>0.057217</td>
<td>0.076290</td>
<td>0.089005</td>
<td>0.098894</td>
<td>0.107135</td>
<td>0.114278</td>
<td>0.120662</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>0.052277</td>
<td>0.069702</td>
<td>0.081320</td>
<td>0.090356</td>
<td>0.097885</td>
<td>0.104354</td>
<td>0.112012</td>
<td>0.115460</td>
</tr>
</tbody>
</table>

### 3. Estimation of $\mu$ and $\lambda$ by BLUE.

Suppose $y_{U(r+1)}, y_{U(r+2)}, y_{U(r+3)}, \ldots, y_{U(n-r-r)}$ is the sequence of upper record values from Weibull distribution for Type II censoring are available. We have the best linear unbiased estimators $*\mu$ and $*\lambda$ expressed as the linear function of upper record statistics, namely,

$$
*\mu = \sum_{i=1}^{n-r-r} a_i y_{U(n)} \quad \text{and} \quad *\lambda = \sum_{i=1}^{n-r-r} b_i y_{U(n)}
$$

where,

$$
a = \frac{1}{\Delta} \left[ \alpha \left( V^{(n)} \right)^{-1} \alpha \right]$$

$$
b = \frac{1}{\Delta} \left[ -\left( V^{(n)} \right)^{-1} \alpha + \left( \left( V^{(n)} \right)^{-1} \alpha \right) \right]$$

The variance and covariance of the above estimators are

$$
\text{Var}(\mu) = \frac{\lambda^2 \alpha ^2 \left( V^{(n)} \right)^{-1} \alpha}{\Delta}
$$

$$
\text{Var}(\lambda) = \frac{\lambda^2 \left( V^{(n)} \right)^{-1} \alpha}{\Delta}
$$

$$
\text{Cov}(\mu, \lambda) = -\frac{\lambda^2 \left( V^{(n)} \right)^{-1} \alpha}{\Delta}
$$

where $\Delta = [(\alpha ^2 \left( V^{(n)} \right)^{-1} \alpha) - (\left( V^{(n)} \right)^{-1} \alpha)^2]$
where \( y^{(n)} \) be the matrix of \( y_U \)

\[
y^{(n)} = \begin{bmatrix}
y^{(n)}_{(r+1)} \\
y^{(n)}_{(r+2)} \\
\vdots \\
y^{(n)}_{(n-\eta-r_2)}
\end{bmatrix}
\]

and \( V^{(n)} = \begin{bmatrix}
V_{(\eta+1),(\eta+1)} & V_{(\eta+1),(\eta+2)} & \cdots & V_{(\eta+1),(n-\eta-r_2)} \\
V_{(\eta+2),(\eta+1)} & V_{(\eta+2),(\eta+2)} & \cdots & V_{(\eta+2),(n-\eta-r_2)} \\
\vdots & \vdots & \ddots & \vdots \\
V_{(n-\eta-r_2+1),(\eta+1)} & V_{(n-\eta-r_2+1),(\eta+2)} & \cdots & V_{(n-\eta-r_2+1),(n-\eta-r_2+1)}
\end{bmatrix}
\]

By using the Tables I and Table II, we computed the coefficients for BLUES of \( \mu \) and \( \lambda \) and the variances and covariances of these estimators for sample size \( n \) up to order 8 and presented in Table III, Table IV and Table V, respectively.

**Table III**

*Showing the coefficients for BLUE of the location parameter based on Upper record values of Weibull distribution from complete samples, singly and doubly censored samples.*

<table>
<thead>
<tr>
<th>( n )</th>
<th>( r_1 )</th>
<th>( r_2 )</th>
<th>( a_1 )</th>
<th>( a_2 )</th>
<th>( a_3 )</th>
<th>( a_4 )</th>
<th>( a_5 )</th>
<th>( a_6 )</th>
<th>( a_7 )</th>
<th>( a_8 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>0</td>
<td>0</td>
<td>4.00000</td>
<td>-3.00000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>0</td>
<td>2.500201</td>
<td>0.749497</td>
<td>-2.249699</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>1</td>
<td>4.00000</td>
<td>-3.00000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td></td>
<td>7.0000007</td>
<td>-6.00000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>0</td>
<td>1.9513342</td>
<td>0.5849651</td>
<td>0.4394632</td>
<td>-1.9758</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>1</td>
<td>2.500201</td>
<td>0.749497</td>
<td>-2.249699</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>2</td>
<td></td>
<td>4.00000</td>
<td>-3.00000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td></td>
<td>3.999609</td>
<td>0.858048</td>
<td>-3.85766</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td></td>
<td>7.0000007</td>
<td>-6.00000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>10.00003</td>
<td>-9.00003</td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>0</td>
<td>1.6543025</td>
<td>0.4959219</td>
<td>0.372568</td>
<td>0.303825</td>
<td>-1.826617</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>1</td>
<td>1.9513342</td>
<td>0.5849651</td>
<td>0.4394632</td>
<td>-1.9758</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>2</td>
<td>2.500201</td>
<td>0.749497</td>
<td>-2.249699</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>3</td>
<td>4.00000</td>
<td>-3.00000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td></td>
<td>2.9614197</td>
<td>0.6353214</td>
<td>0.5180976</td>
<td>-3.114839</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td></td>
<td>3.999609</td>
<td>0.858048</td>
<td>-3.85766</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td></td>
<td>7.0000007</td>
<td>-6.00000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td></td>
<td>5.5008790</td>
<td>0.8980229</td>
<td>-5.398902</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>10.00003</td>
<td>-9.00003</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td></td>
<td>12.99987</td>
<td>-11.99987</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>0</td>
<td>0</td>
<td>1.4631735</td>
<td>0.4386258</td>
<td>0.3295234</td>
<td>0.2687230</td>
<td>0.2330143</td>
<td>-1.7330601</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>0</td>
<td>1</td>
<td>1.6543025</td>
<td>0.4959219</td>
<td>0.372568</td>
<td>0.303825</td>
<td>-1.826617</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>2</td>
<td>1.9513342</td>
<td>0.5849651</td>
<td>0.4394632</td>
<td>-1.9758</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>3</td>
<td>2.500201</td>
<td>0.749497</td>
<td>-2.249699</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>n</td>
<td>r₁</td>
<td>r₂</td>
<td>a₁</td>
<td>a₂</td>
<td>a₃</td>
<td>a₄</td>
<td>a₅</td>
<td>a₆</td>
<td>a₇</td>
<td>a₈</td>
</tr>
<tr>
<td>----</td>
<td>-----</td>
<td>-----</td>
<td>------</td>
<td>--------</td>
<td>--------</td>
<td>--------</td>
<td>--------</td>
<td>--------</td>
<td>--------</td>
<td>--------</td>
</tr>
<tr>
<td>0</td>
<td>4</td>
<td>0.00000</td>
<td>-3.00000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>2.42193296</td>
<td>0.5195838</td>
<td>0.4237153</td>
<td>0.3674107</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>2.9614197</td>
<td>0.6353214</td>
<td>0.5180976</td>
<td>-3.114839</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>3.999609</td>
<td>0.858048</td>
<td>-3.85766</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>3</td>
<td>7.0000007</td>
<td>-6.00000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>3.9863033</td>
<td>0.6478828</td>
<td>0.5617824</td>
<td>-4.178296</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>5.008790</td>
<td>0.8980229</td>
<td>-5.398902</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>10.00003</td>
<td>-9.00003</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>6.9961623</td>
<td>0.9314325</td>
<td>-6.9275947</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>12.9987</td>
<td>-11.9987</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>16.00037</td>
<td>-15.00037</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>1.328506</td>
<td>0.398255</td>
<td>0.299195</td>
<td>0.211568</td>
<td>-1.656603</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>1.4631735</td>
<td>0.4386258</td>
<td>0.3295234</td>
<td>0.2687230</td>
<td>0.2330143</td>
<td>-1.7330601</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>2</td>
<td>1.6543025</td>
<td>0.4959219</td>
<td>0.372568</td>
<td>0.303825</td>
<td>-1.826617</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>3</td>
<td>1.9513342</td>
<td>0.5849651</td>
<td>0.4394632</td>
<td>-1.9758</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>2.500201</td>
<td>0.749497</td>
<td>-2.249699</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>4.00000</td>
<td>-3.00000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>2.08817159</td>
<td>0.447981</td>
<td>0.3653241</td>
<td>0.3167786</td>
<td>0.262158</td>
<td>-2.480413</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>2.42193296</td>
<td>0.5195838</td>
<td>0.4237153</td>
<td>0.3674107</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>2</td>
<td>2.9614197</td>
<td>0.6353214</td>
<td>0.5180976</td>
<td>-3.114839</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>3</td>
<td>3.999609</td>
<td>0.858048</td>
<td>-3.85766</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>4</td>
<td>7.0000007</td>
<td>-6.00000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>0</td>
<td>3.1892135</td>
<td>0.5206424</td>
<td>0.4514516</td>
<td>0.3736099</td>
<td>-3.534917</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>3.9686303</td>
<td>0.6478828</td>
<td>0.5617824</td>
<td>-4.178296</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>2</td>
<td>5.5008790</td>
<td>0.8980229</td>
<td>-5.398902</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>3</td>
<td>10.00003</td>
<td>-9.00003</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>4</td>
<td>6.9961623</td>
<td>0.9314325</td>
<td>-6.9275947</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>5</td>
<td>12.99987</td>
<td>-11.99987</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>0</td>
<td>8.5227025</td>
<td>0.8890491</td>
<td>-8.411752</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>0</td>
<td>18.99907</td>
<td>-17.99907</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>0</td>
<td>1.220929</td>
<td>0.366006</td>
<td>0.274967</td>
<td>0.224233</td>
<td>0.198724</td>
<td>-0.016646</td>
<td>0.618563</td>
<td>-1.886776</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>1</td>
<td>1.328506</td>
<td>0.398255</td>
<td>0.299195</td>
<td>0.243990</td>
<td>0.211568</td>
<td>0.175088</td>
<td>-1.656603</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>2</td>
<td>1.4631735</td>
<td>0.4386258</td>
<td>0.3295234</td>
<td>0.2687230</td>
<td>0.2330143</td>
<td>-1.7330601</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>3</td>
<td>1.6543025</td>
<td>0.4959219</td>
<td>0.372568</td>
<td>0.303825</td>
<td>-1.826617</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>4</td>
<td>1.9513342</td>
<td>0.5849651</td>
<td>0.4394632</td>
<td>-1.9758</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>5</td>
<td>2.500201</td>
<td>0.749497</td>
<td>-2.249699</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>6</td>
<td>4.0000000</td>
<td>-3.00000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>0</td>
<td>1.844793</td>
<td>0.395768</td>
<td>0.322745</td>
<td>0.286029</td>
<td>-0.023959</td>
<td>0.890316</td>
<td>-2.715692</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>1</td>
<td>2.0881717</td>
<td>0.447981</td>
<td>0.365324</td>
<td>0.316779</td>
<td>0.262158</td>
<td>-2.480413</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>2</td>
<td>2.421933</td>
<td>0.519584</td>
<td>0.423715</td>
<td>0.367411</td>
<td>-2.732643</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>3</td>
<td>2.961420</td>
<td>0.635321</td>
<td>0.518098</td>
<td>-3.114839</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Estimation of the Scale and location parameters based upon …

<table>
<thead>
<tr>
<th>n</th>
<th>r&lt;sub&gt;1&lt;/sub&gt;</th>
<th>r&lt;sub&gt;2&lt;/sub&gt;</th>
<th>a&lt;sub&gt;1&lt;/sub&gt;</th>
<th>a&lt;sub&gt;2&lt;/sub&gt;</th>
<th>a&lt;sub&gt;3&lt;/sub&gt;</th>
<th>a&lt;sub&gt;4&lt;/sub&gt;</th>
<th>a&lt;sub&gt;5&lt;/sub&gt;</th>
<th>a&lt;sub&gt;6&lt;/sub&gt;</th>
<th>a&lt;sub&gt;7&lt;/sub&gt;</th>
<th>a&lt;sub&gt;8&lt;/sub&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>4</td>
<td>3.999609</td>
<td>0.858048</td>
<td>-3.85766</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>5</td>
<td>7.0000007</td>
<td>-6.00000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>2.684492</td>
<td>0.438246</td>
<td>0.388385</td>
<td>-0.032533</td>
<td>1.208917</td>
<td>-3.687507</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>3.189212</td>
<td>0.520642</td>
<td>0.451451</td>
<td>0.373610</td>
<td>-3.534916</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>3.968625</td>
<td>0.647882</td>
<td>0.561782</td>
<td>-4.178290</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>5.500879</td>
<td>0.898023</td>
<td>-5.398902</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>4</td>
<td>10.00003</td>
<td>-9.00003</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>3.901694</td>
<td>0.530906</td>
<td>-0.004472</td>
<td>1.652539</td>
<td>-5.040667</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>4.978766</td>
<td>0.662847</td>
<td>0.548555</td>
<td>-5.190168</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>6.996162</td>
<td>0.931433</td>
<td>-6.927595</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>12.99987</td>
<td>-11.99987</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>5.904648</td>
<td>-0.063543</td>
<td>2.361220</td>
<td>-7.20233</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>8.522703</td>
<td>0.889049</td>
<td>-8.411752</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>2</td>
<td>16.00003</td>
<td>-15.00003</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>8.690069</td>
<td>3.705238</td>
<td>-11.39531</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>18.99907</td>
<td>-17.99907</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>0</td>
<td>22.00108</td>
<td>-21.00108</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table IV

Showing the coefficients for BLUE of scale parameter Based on Upper record values of Weibull distribution from complete samples, singly and doubly censored samples

<table>
<thead>
<tr>
<th>n</th>
<th>r&lt;sub&gt;1&lt;/sub&gt;</th>
<th>r&lt;sub&gt;2&lt;/sub&gt;</th>
<th>b&lt;sub&gt;1&lt;/sub&gt;</th>
<th>b&lt;sub&gt;2&lt;/sub&gt;</th>
<th>b&lt;sub&gt;3&lt;/sub&gt;</th>
<th>b&lt;sub&gt;4&lt;/sub&gt;</th>
<th>b&lt;sub&gt;5&lt;/sub&gt;</th>
<th>b&lt;sub&gt;6&lt;/sub&gt;</th>
<th>b&lt;sub&gt;7&lt;/sub&gt;</th>
<th>b&lt;sub&gt;8&lt;/sub&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>0</td>
<td>0</td>
<td>-2.329373</td>
<td>2.329373</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>0</td>
<td>-1.246818</td>
<td>-0.373764</td>
<td>1.620582</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>1</td>
<td>-2.329373</td>
<td>2.329373</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>0</td>
<td>-0.876610</td>
<td>-0.262788</td>
<td>-0.197423</td>
<td>1.336821</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>1</td>
<td>-1.246818</td>
<td>-0.373764</td>
<td>1.620582</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>2</td>
<td>-2.329373</td>
<td>2.329373</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>0</td>
<td>-0.686005</td>
<td>-0.205649</td>
<td>-0.154496</td>
<td>-0.125990</td>
<td>1.172140</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>1</td>
<td>-0.876610</td>
<td>-0.262788</td>
<td>-0.197423</td>
<td>1.336821</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>0</td>
<td>0</td>
<td>-0.568827</td>
<td>-0.170521</td>
<td>-0.128107</td>
<td>-0.104469</td>
<td>-0.090587</td>
<td>1.0625115</td>
<td></td>
<td></td>
</tr>
<tr>
<td>n</td>
<td>r_1</td>
<td>r_2</td>
<td>b_1</td>
<td>b_2</td>
<td>b_3</td>
<td>b_4</td>
<td>b_5</td>
<td>b_6</td>
<td>b_7</td>
<td>b_8</td>
</tr>
<tr>
<td>---</td>
<td>-----</td>
<td>-----</td>
<td>---------</td>
<td>---------</td>
<td>---------</td>
<td>---------</td>
<td>---------</td>
<td>---------</td>
<td>---------</td>
<td>---------</td>
</tr>
<tr>
<td>6</td>
<td>0</td>
<td>1</td>
<td>-0.686005</td>
<td>-0.205649</td>
<td>-0.154496</td>
<td>-0.125990</td>
<td>1.172140</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>2</td>
<td>-0.876610</td>
<td>-0.262788</td>
<td>-0.197423</td>
<td>1.336821</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>3</td>
<td>-1.246818</td>
<td>-0.373764</td>
<td>1.620582</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>4</td>
<td>-2.329373</td>
<td>2.329373</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>-0.941557</td>
<td>-0.201995</td>
<td>-0.164724</td>
<td>-0.14283</td>
<td>1.451112</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>-1.228040</td>
<td>-0.263455</td>
<td>-0.214844</td>
<td>1.706339</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>-1.79677</td>
<td>-0.385466</td>
<td>2.182237</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>3</td>
<td>-3.494060</td>
<td>3.494060</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>-1.542855</td>
<td>-0.251873</td>
<td>-0.218400</td>
<td>2.013128</td>
<td>-1.542855</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>-2.281101</td>
<td>-0.372391</td>
<td>2.653493</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>-4.492376</td>
<td>4.492376</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>-2.719846</td>
<td>-0.362106</td>
<td>3.081952</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>-5.390777</td>
<td>5.390777</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>-6.220347</td>
<td>6.220347</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>0</td>
<td>-0.489335</td>
<td>-0.146691</td>
<td>-0.110204</td>
<td>-0.08987</td>
<td>-0.077928</td>
<td>-0.063835</td>
<td>0.977863</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>0</td>
<td>-0.568827</td>
<td>-0.170521</td>
<td>-0.128107</td>
<td>-0.104469</td>
<td>-0.090587</td>
<td>1.0625115</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>2</td>
<td>-0.686005</td>
<td>-0.205649</td>
<td>-0.154496</td>
<td>-0.125990</td>
<td>1.172140</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>3</td>
<td>-0.876610</td>
<td>-0.262788</td>
<td>-0.197423</td>
<td>1.336821</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>4</td>
<td>-1.246818</td>
<td>-0.373764</td>
<td>1.620582</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>5</td>
<td>-2.329373</td>
<td>2.329373</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>-0.769147</td>
<td>-0.165007</td>
<td>-0.134562</td>
<td>-0.116681</td>
<td>-0.095906</td>
<td>1.28130143</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>-0.941557</td>
<td>-0.201995</td>
<td>-0.164724</td>
<td>-0.14283</td>
<td>1.451112</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>-1.228040</td>
<td>-0.263455</td>
<td>-0.214844</td>
<td>1.706339</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>3</td>
<td>-1.79677</td>
<td>-0.385466</td>
<td>2.182237</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>4</td>
<td>-3.494060</td>
<td>3.494060</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>-1.174699</td>
<td>-0.191771</td>
<td>-0.166285</td>
<td>-0.136957</td>
<td>1.669712</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>-1.542855</td>
<td>-0.251873</td>
<td>-0.218400</td>
<td>2.013128</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>-2.281101</td>
<td>-0.372391</td>
<td>2.653493</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>-4.492376</td>
<td>4.492376</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>-1.833854</td>
<td>-0.244150</td>
<td>-0.201396</td>
<td>2.279399</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>-2.719846</td>
<td>-0.362106</td>
<td>3.081952</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>-5.390777</td>
<td>5.390777</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>-3.139209</td>
<td>-0.326812</td>
<td>3.466020</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>-6.220347</td>
<td>6.220347</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>-6.997366</td>
<td>6.997366</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>0</td>
<td>-0.428396</td>
<td>-0.128423</td>
<td>-0.0964796</td>
<td>0.0786781</td>
<td>-0.070652</td>
<td>0.0447779</td>
<td>-0.31096</td>
<td>1.0688102</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>0</td>
<td>-0.489335</td>
<td>-0.146691</td>
<td>-0.110204</td>
<td>-0.08987</td>
<td>-0.077928</td>
<td>-0.063835</td>
<td>0.977863</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>2</td>
<td>-0.568827</td>
<td>-0.170521</td>
<td>-0.128107</td>
<td>-0.104469</td>
<td>-0.090587</td>
<td>1.062512</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>3</td>
<td>-0.686005</td>
<td>-0.205649</td>
<td>-0.154496</td>
<td>-0.125990</td>
<td>1.172140</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>4</td>
<td>-0.876610</td>
<td>-0.262788</td>
<td>-0.197423</td>
<td>1.336821</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>5</td>
<td>-1.246818</td>
<td>-0.373764</td>
<td>1.620582</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>6</td>
<td>-2.329373</td>
<td>2.329373</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Estimation of the Scale and location parameters based upon …

<table>
<thead>
<tr>
<th>n</th>
<th>r_1</th>
<th>r_2</th>
<th>b_1</th>
<th>b_2</th>
<th>b_3</th>
<th>b_4</th>
<th>b_5</th>
<th>b_6</th>
<th>b_7</th>
<th>b_8</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td></td>
<td>-0.647295</td>
<td>-0.138866</td>
<td>-0.113244</td>
<td>-0.101285</td>
<td>0.047344</td>
<td>-0.4063116</td>
<td>1.3596578</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td></td>
<td>-0.769147</td>
<td>-0.165007</td>
<td>-0.134562</td>
<td>-0.116681</td>
<td>-0.095906</td>
<td>1.2813014</td>
<td>1.706339</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td></td>
<td>-0.941557</td>
<td>-0.201995</td>
<td>-0.164724</td>
<td>-0.14283</td>
<td>1.451112</td>
<td>-0.095906</td>
<td>1.2813014</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>3</td>
<td></td>
<td>-1.228040</td>
<td>-0.263455</td>
<td>-0.214844</td>
<td>1.706339</td>
<td>-0.095906</td>
<td>1.2813014</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>4</td>
<td></td>
<td>-1.79677</td>
<td>-0.385466</td>
<td>2.182237</td>
<td>1.706339</td>
<td>-0.095906</td>
<td>1.2813014</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>5</td>
<td></td>
<td>-3.494060</td>
<td>3.494060</td>
<td>3.494060</td>
<td>3.494060</td>
<td>3.494060</td>
<td>3.494060</td>
<td>3.494060</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td></td>
<td>-0.941926</td>
<td>-0.153770</td>
<td>-0.137110</td>
<td>0.050352</td>
<td>-0.518101</td>
<td>1.700645</td>
<td>1.700645</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td></td>
<td>-1.174699</td>
<td>-0.191771</td>
<td>-0.166285</td>
<td>-0.134562</td>
<td>1.669712</td>
<td>1.700645</td>
<td>1.700645</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td></td>
<td>-1.542855</td>
<td>-0.251873</td>
<td>-0.218400</td>
<td>2.013128</td>
<td>1.700645</td>
<td>1.700645</td>
<td>1.700645</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td></td>
<td>-2.281101</td>
<td>-0.372391</td>
<td>2.653493</td>
<td>1.700645</td>
<td>1.700645</td>
<td>1.700645</td>
<td>1.700645</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>4</td>
<td></td>
<td>-4.492376</td>
<td>4.492376</td>
<td>4.492376</td>
<td>4.492376</td>
<td>4.492376</td>
<td>4.492376</td>
<td>4.492376</td>
<td></td>
</tr>
</tbody>
</table>

Table V
Variances and covariances of BLUEs for location and scale based on the Upper record values from the Weibull distribution from complete samples, singly and doubly censored samples.

Each value may be multiplied by $\lambda^2$

<table>
<thead>
<tr>
<th>n</th>
<th>r_1</th>
<th>r_2</th>
<th>Var($\mu$)</th>
<th>Var($\lambda$)</th>
<th>Cov($\mu$, $\lambda$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>0</td>
<td>0</td>
<td>1.877692</td>
<td>0.698109</td>
<td>1.093461</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>0</td>
<td>1.173652</td>
<td>0.332094</td>
<td>0.585830</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>1</td>
<td>1.173652</td>
<td>0.332094</td>
<td>0.585830</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>0</td>
<td>0.916001</td>
<td>0.214143</td>
<td>0.41150</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>1</td>
<td>1.173652</td>
<td>0.332094</td>
<td>0.585830</td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>0</td>
<td>0.776567</td>
<td>0.156726</td>
<td>0.322026</td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>1</td>
<td>0.916001</td>
<td>0.214143</td>
<td>0.41150</td>
</tr>
<tr>
<td>6</td>
<td>0</td>
<td>0</td>
<td>1.877692</td>
<td>0.698109</td>
<td>1.093461</td>
</tr>
<tr>
<td>n</td>
<td>r₁</td>
<td>r₂</td>
<td>( \text{Var}(\mu) )</td>
<td>( \text{Var}(\lambda) )</td>
<td>( \text{Cov}(\mu, \lambda) )</td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1.324241</td>
<td>0.250903</td>
<td>0.549136</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1.788482</td>
<td>0.390219</td>
<td>0.803452</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>2</td>
<td>3.130157</td>
<td>0.819562</td>
<td>1.562423</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>0</td>
<td>2.295287</td>
<td>0.417883</td>
<td>0.951809</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>1</td>
<td>4.172593</td>
<td>0.871365</td>
<td>1.8744809</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>0</td>
<td>5.101619</td>
<td>0.900487</td>
<td>2.115536</td>
</tr>
<tr>
<td>6</td>
<td>0</td>
<td>0</td>
<td>0.686847</td>
<td>0.123002</td>
<td>0.267020</td>
</tr>
<tr>
<td>6</td>
<td>0</td>
<td>1</td>
<td>0.776567</td>
<td>0.156726</td>
<td>0.322026</td>
</tr>
<tr>
<td>0</td>
<td>2</td>
<td>0.916001</td>
<td>0.214143</td>
<td>0.411501</td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>3</td>
<td>1.173652</td>
<td>0.332094</td>
<td>0.585830</td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>4</td>
<td>1.877692</td>
<td>0.698109</td>
<td>1.093461</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1.083002</td>
<td>0.182876</td>
<td>0.421031</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1.324241</td>
<td>0.250903</td>
<td>0.549136</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>1.788482</td>
<td>0.390219</td>
<td>0.803452</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>3</td>
<td>3.130157</td>
<td>0.819562</td>
<td>1.562423</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>0</td>
<td>1.655944</td>
<td>0.269468</td>
<td>0.643769</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>2.295287</td>
<td>0.417883</td>
<td>0.951809</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>4.172593</td>
<td>0.871365</td>
<td>1.8744809</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>0</td>
<td>2.745547</td>
<td>0.434147</td>
<td>1.067366</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>5.101619</td>
<td>0.900487</td>
<td>2.115536</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>0</td>
<td>5.944956</td>
<td>0.917693</td>
<td>2.311767</td>
</tr>
<tr>
<td>7</td>
<td>0</td>
<td>0</td>
<td>0.623631</td>
<td>0.100976</td>
<td>0.229705</td>
</tr>
<tr>
<td>7</td>
<td>0</td>
<td>1</td>
<td>0.686847</td>
<td>0.123002</td>
<td>0.267020</td>
</tr>
<tr>
<td>0</td>
<td>2</td>
<td>0.776567</td>
<td>0.156726</td>
<td>0.322026</td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>3</td>
<td>0.916001</td>
<td>0.214143</td>
<td>0.411501</td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>4</td>
<td>1.173652</td>
<td>0.332094</td>
<td>0.585830</td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>5</td>
<td>1.877692</td>
<td>0.698109</td>
<td>1.093461</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0.933756</td>
<td>0.143050</td>
<td>0.343935</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1.083002</td>
<td>0.182876</td>
<td>0.421031</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>1.324241</td>
<td>0.250903</td>
<td>0.549136</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>3</td>
<td>1.788482</td>
<td>0.390219</td>
<td>0.803452</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>4</td>
<td>3.130157</td>
<td>0.819562</td>
<td>1.562423</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>0</td>
<td>1.330726</td>
<td>0.196908</td>
<td>0.490153</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>1.655944</td>
<td>0.269468</td>
<td>0.643769</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>2.295287</td>
<td>0.417883</td>
<td>0.951809</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>4.172593</td>
<td>0.871365</td>
<td>1.8744809</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>0</td>
<td>1.953848</td>
<td>0.281447</td>
<td>0.719670</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>2.745547</td>
<td>0.434147</td>
<td>1.067366</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>5.101619</td>
<td>0.900487</td>
<td>2.115536</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>0</td>
<td>3.166619</td>
<td>0.445984</td>
<td>1.166396</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>5.944956</td>
<td>0.917693</td>
<td>2.311767</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>0</td>
<td>6.775784</td>
<td>0.935642</td>
<td>2.495759</td>
</tr>
<tr>
<td>8</td>
<td>0</td>
<td>0</td>
<td>0.573132</td>
<td>0.084771</td>
<td>0.201099</td>
</tr>
</tbody>
</table>
Lloyd’s procedure requires full knowledge of the expectations and the variance covariance matrix of the record values. The covariance be specially may be difficult to determine. Gupta (1952) has proposed a very simple method applicable when only the expectations are known. The coefficients of these linear estimates are obtained by assuming the variance matrix to be a unit matrix. Let the linear estimates be

\[ \mu = \sum_{i=r_1+1}^{n-r_2} b_i x_{(i)}^n \] \[ \lambda = \sum_{i=r_1+1}^{n-r_2} c_i x_{(i)}^n \]

where

\[ b_i = \frac{1}{n-r_1-r_2} \sum_{i=r_1+1}^{n-r_2} (\bar{\mu}_i - \bar{\mu}_k)^2 \]

\[ c_i = \frac{1}{n-r_1-r_2} \sum_{i=r_1+1}^{n-r_2} (\mu_i - \bar{\mu}_k)^2 \]

and \( \bar{\mu}_k = \frac{1}{n-r_1-r_2} \sum_{j=r_1+1}^{n-r_2} \mu_j \)}

<table>
<thead>
<tr>
<th>n</th>
<th>r_1</th>
<th>r_2</th>
<th>Var(\mu)</th>
<th>Var(\lambda)</th>
<th>Cov(\mu, \lambda)</th>
</tr>
</thead>
<tbody>
<tr>
<td>8</td>
<td>0</td>
<td>1</td>
<td>0.623631</td>
<td>0.100976</td>
<td>0.229705</td>
</tr>
<tr>
<td>0</td>
<td>2</td>
<td>0.686847</td>
<td>0.123002</td>
<td>0.267020</td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>3</td>
<td>0.776567</td>
<td>0.156726</td>
<td>0.322026</td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>4</td>
<td>0.916001</td>
<td>0.214143</td>
<td>0.411501</td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>5</td>
<td>1.173652</td>
<td>0.332094</td>
<td>0.585830</td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>6</td>
<td>1.877692</td>
<td>0.698109</td>
<td>1.093461</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>0.824926</td>
<td>0.115770</td>
<td>0.289447</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>0.933756</td>
<td>0.143050</td>
<td>0.421031</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>1.083002</td>
<td>0.182876</td>
<td>0.549136</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>3</td>
<td>1.324241</td>
<td>0.250903</td>
<td>0.803452</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>4</td>
<td>1.788482</td>
<td>0.390219</td>
<td>1.166396</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>5</td>
<td>3.130157</td>
<td>0.819562</td>
<td>1.562423</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>1.120127</td>
<td>0.152114</td>
<td>0.393026</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>1.330726</td>
<td>0.196908</td>
<td>0.490153</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>1.655944</td>
<td>0.269468</td>
<td>0.643769</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>2.295287</td>
<td>0.417883</td>
<td>0.951809</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>4</td>
<td>4.172593</td>
<td>0.871365</td>
<td>1.874480</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>1.531166</td>
<td>0.202719</td>
<td>0.537250</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>1.953848</td>
<td>0.281447</td>
<td>0.719670</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>2.745547</td>
<td>0.43147</td>
<td>1.067366</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>5.101619</td>
<td>0.900487</td>
<td>2.115536</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>2.187797</td>
<td>0.283559</td>
<td>0.767647</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>3.166619</td>
<td>0.445984</td>
<td>1.166396</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>2</td>
<td>5.944956</td>
<td>0.917693</td>
<td>2.311176</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>3.472993</td>
<td>0.441927</td>
<td>1.218794</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>6.775784</td>
<td>0.935642</td>
<td>2.495759</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>0</td>
<td>5.819889</td>
<td>0.723788</td>
<td>2.032119</td>
<td></td>
</tr>
</tbody>
</table>

4. AN ALTERNATIVE LINEAR ESTIMATE (GUPTA 1952)

Lloyd’s procedure requires full knowledge of the expectations and the variance covariance matrix of the record values. The covariance be specially may be difficult to determine. Gupta (1952) has proposed a very simple method applicable when only the expectations are known. The coefficients of these linear estimates are obtained by assuming the variance matrix to be a unit matrix. Let the linear estimates be

\[ \mu = \sum_{i=r_1+1}^{n-r_2} b_i x_{(i)}^n \]

\[ \lambda = \sum_{i=r_1+1}^{n-r_2} c_i x_{(i)}^n \]

where

\[ b_i = \frac{1}{n-r_1-r_2} \sum_{i=r_1+1}^{n-r_2} (\bar{\mu}_i - \bar{\mu}_k)^2 \]

\[ c_i = \frac{1}{n-r_1-r_2} \sum_{i=r_1+1}^{n-r_2} (\mu_i - \bar{\mu}_k)^2 \]

and \( \bar{\mu}_k = \frac{1}{n-r_1-r_2} \sum_{j=r_1+1}^{n-r_2} \mu_j \)
In matrix notation the variance of the estimates can also be written as
\[
V(\mu) = \lambda^2 b'Vb
\]
\[
V(\lambda) = \lambda^2 c'Vc
\]

By making use of Table I and II, we determine the coefficients of the estimates and variances of the estimates are presented in Table VI, VII and VIII respectively.

Table VI

Showing the coefficients for alternative linear estimate of location parameters based on Upper Records values of Weibull distribution complete samples, singly and doubly censored samples.

<table>
<thead>
<tr>
<th>n</th>
<th>r_1</th>
<th>r_2</th>
<th>b_1</th>
<th>b_2</th>
<th>b_3</th>
<th>b_4</th>
<th>b_5</th>
<th>b_6</th>
<th>b_7</th>
<th>b_8</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>0</td>
<td>0</td>
<td>4</td>
<td>-3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>0</td>
<td>2.789474</td>
<td>0.026316</td>
<td>-1.815789</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>7</td>
<td>-6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>0</td>
<td>2.16558</td>
<td>0.586322</td>
<td>-0.466514</td>
<td>-1.285384</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>1</td>
<td>2.789474</td>
<td>0.026316</td>
<td>-1.815789</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>2</td>
<td>4</td>
<td>-3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>4.37047</td>
<td>0.010359</td>
<td>-3.38083</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>7</td>
<td>-6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>10</td>
<td>-9.00003</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>0</td>
<td>1.78124</td>
<td>0.72273</td>
<td>0.017047</td>
<td>-0.531813</td>
<td>-0.989203</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>1</td>
<td>2.16558</td>
<td>0.586322</td>
<td>-0.466514</td>
<td>-1.285384</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>2</td>
<td>4.37047</td>
<td>0.010359</td>
<td>-3.38083</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>1</td>
<td>2.789474</td>
<td>0.026316</td>
<td>-1.815789</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>3</td>
<td>4</td>
<td>-3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>3.14927</td>
<td>0.920651</td>
<td>-0.812715</td>
<td>-2.25721</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>4.37047</td>
<td>0.010359</td>
<td>-3.38083</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>7</td>
<td>-6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>5.90657</td>
<td>0.00552</td>
<td>-4.91209</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>10</td>
<td>-9.00003</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>12.9999</td>
<td>-11.9999</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>0</td>
<td>0</td>
<td>1.519056</td>
<td>0.743374</td>
<td>0.226252</td>
<td>-0.17595</td>
<td>-0.511128</td>
<td>-0.801602</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>0</td>
<td>1</td>
<td>1.78124</td>
<td>0.72273</td>
<td>0.017047</td>
<td>-0.531813</td>
<td>-0.989203</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>2</td>
<td>2.16558</td>
<td>0.586322</td>
<td>-0.466514</td>
<td>-1.285384</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>3</td>
<td>2.789474</td>
<td>0.026316</td>
<td>-1.815789</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>4</td>
<td>4</td>
<td>-3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>2.45388</td>
<td>1.07811</td>
<td>0.00807315</td>
<td>-0.883638</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>3.14927</td>
<td>0.920651</td>
<td>-0.812715</td>
<td>-2.25721</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>4.37047</td>
<td>0.010359</td>
<td>-3.38083</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>3</td>
<td>7</td>
<td>-6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>4.08818</td>
<td>1.23923</td>
<td>-1.13494</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>5.90657</td>
<td>0.00552</td>
<td>-4.91209</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>n</td>
<td>r₁</td>
<td>r₂</td>
<td>b₁</td>
<td>b₂</td>
<td>b₃</td>
<td>b₄</td>
<td>b₅</td>
<td>b₆</td>
<td>b₇</td>
<td>b₈</td>
</tr>
<tr>
<td>----</td>
<td>----</td>
<td>----</td>
<td>-----</td>
<td>-------</td>
<td>-------</td>
<td>-------</td>
<td>-------</td>
<td>-------</td>
<td>-------</td>
<td>-------</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>10</td>
<td>-9.00003</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>7.42707</td>
<td>0.003313</td>
<td>-6.43038</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>12.9999</td>
<td>-11.9999</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>16.0004</td>
<td>-15.0004</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>0</td>
<td>0.1327923</td>
<td>0.3250006</td>
<td>0.0129886</td>
<td>-0.24703</td>
<td>-0.472371</td>
<td>-0.672689</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>0</td>
<td>1.519056</td>
<td>0.226252</td>
<td>-0.17595</td>
<td>-0.511128</td>
<td>-0.801602</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>2</td>
<td>1.78124</td>
<td>0.017047</td>
<td>-0.531813</td>
<td>-0.989203</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>3</td>
<td>2.16558</td>
<td>-0.466514</td>
<td>-1.285384</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>4</td>
<td>2.789474</td>
<td>0.026316</td>
<td>-1.815789</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>5</td>
<td>7</td>
<td>-6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>2.00792</td>
<td>1.06194</td>
<td>0.326173</td>
<td>-0.286972</td>
<td>-0.818346</td>
<td>-1.29071</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>2.45388</td>
<td>1.07811</td>
<td>0.00807315</td>
<td>-0.883638</td>
<td>-1.65643</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>3.14927</td>
<td>0.920651</td>
<td>-0.812715</td>
<td>-2.25721</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>3</td>
<td>4.37047</td>
<td>0.010359</td>
<td>-3.38083</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>4</td>
<td>7</td>
<td>-6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>3.08886</td>
<td>1.40664</td>
<td>0.004773</td>
<td>-1.21014</td>
<td>-2.29014</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>4.08818</td>
<td>1.23923</td>
<td>-1.13494</td>
<td>-3.19247</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>5.90657</td>
<td>0.00552</td>
<td>-4.91209</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>10</td>
<td>-9.00003</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>5.01085</td>
<td>1.55073</td>
<td>-1.44795</td>
<td>-4.11363</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>7.42707</td>
<td>0.003313</td>
<td>-6.43038</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>12.9999</td>
<td>-11.9999</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>8.93991</td>
<td>0.002517</td>
<td>-7.94243</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>16.0004</td>
<td>-15.0004</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>18.9991</td>
<td>-17.9991</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>0</td>
<td>0.1327923</td>
<td>0.3250006</td>
<td>0.0129886</td>
<td>-0.24703</td>
<td>-0.472371</td>
<td>-0.672689</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>0</td>
<td>1.519056</td>
<td>0.226252</td>
<td>-0.17595</td>
<td>-0.511128</td>
<td>-0.801602</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>2</td>
<td>1.78124</td>
<td>0.017047</td>
<td>-0.531813</td>
<td>-0.989203</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>3</td>
<td>2.16558</td>
<td>-0.466514</td>
<td>-1.285384</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>4</td>
<td>2.789474</td>
<td>0.026316</td>
<td>-1.815789</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>6</td>
<td>7</td>
<td>-6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>1.69876</td>
<td>1.00136</td>
<td>0.458941</td>
<td>0.00691878</td>
<td>-0.38482</td>
<td>-0.733058</td>
<td>-1.0481</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>2.00792</td>
<td>1.06194</td>
<td>0.326173</td>
<td>-0.286972</td>
<td>-0.818346</td>
<td>-1.29071</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>2.45388</td>
<td>1.07811</td>
<td>0.00807315</td>
<td>-0.883638</td>
<td>-1.65643</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>3</td>
<td>3.14927</td>
<td>0.920651</td>
<td>-0.812715</td>
<td>-2.25721</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>4</td>
<td>4.37047</td>
<td>0.010359</td>
<td>-3.38083</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>5</td>
<td>7</td>
<td>-6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>2.46656</td>
<td>1.35054</td>
<td>0.420516</td>
<td>-0.38548</td>
<td>-1.10197</td>
<td>-1.75017</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>3.08886</td>
<td>1.40664</td>
<td>0.004773</td>
<td>-1.21014</td>
<td>-2.29014</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>4.08818</td>
<td>1.23923</td>
<td>-1.13494</td>
<td>-3.19247</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>5.90657</td>
<td>0.00552</td>
<td>-4.91209</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>4</td>
<td>10</td>
<td>-9.00003</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>n</td>
<td>r1</td>
<td>r2</td>
<td>b1</td>
<td>b2</td>
<td>b3</td>
<td>b4</td>
<td>b5</td>
<td>b6</td>
<td>b7</td>
<td>b8</td>
</tr>
<tr>
<td>----</td>
<td>----</td>
<td>----</td>
<td>-------</td>
<td>--------</td>
<td>--------</td>
<td>--------</td>
<td>--------</td>
<td>--------</td>
<td>--------</td>
<td>--------</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>0</td>
<td>3.70996</td>
<td>1.72416</td>
<td>0.003203</td>
<td>-1.52665</td>
<td>-2.91067</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>2</td>
<td>5.01085</td>
<td>1.55073</td>
<td>-1.44795</td>
<td>-4.11363</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>2</td>
<td>7.42707</td>
<td>0.003313</td>
<td>-6.43038</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>0</td>
<td>12.9999</td>
<td>-11.9999</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>0</td>
<td>5.92586</td>
<td>1.85869</td>
<td>-1.75684</td>
<td>-2.91067</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>2</td>
<td>0</td>
<td>8.93991</td>
<td>0.002517</td>
<td>-7.94243</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>0</td>
<td>16.0004</td>
<td>-15.0004</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>0</td>
<td>10.4493</td>
<td>0.001335</td>
<td>-9.45067</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>0</td>
<td>0</td>
<td>22.0011</td>
<td>-21.0011</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Tables VII

Showing the coefficients for alternative linear estimate of scale parameter based on
Upper record values of Weibull distribution from complete samples, singly and
doubly censored samples.

<table>
<thead>
<tr>
<th>n</th>
<th>r1</th>
<th>r2</th>
<th>c1</th>
<th>c2</th>
<th>c3</th>
<th>c4</th>
<th>c5</th>
<th>c6</th>
<th>c7</th>
<th>c8</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>0</td>
<td>0</td>
<td>-2.32937</td>
<td>2.32937</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>0</td>
<td>-1.471183</td>
<td>0.183898</td>
<td>1.287285</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>1</td>
<td>-2.32937</td>
<td>2.32937</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>-3.49406</td>
<td>3.49406</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>0</td>
<td>-1.05913</td>
<td>-0.185228</td>
<td>0.396165</td>
<td>0.848923</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>1</td>
<td>-1.471183</td>
<td>0.183898</td>
<td>1.287285</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>2</td>
<td>0</td>
<td>-2.32937</td>
<td>2.32937</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>-2.0367</td>
<td>0.162937</td>
<td>1.87376</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>0</td>
<td>-3.49406</td>
<td>3.49406</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>0</td>
<td>-4.49238</td>
<td>4.49238</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>0</td>
<td>-0.819637</td>
<td>-0.270955</td>
<td>0.094834</td>
<td>0.379335</td>
<td>0.616423</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>1</td>
<td>-1.05913</td>
<td>-0.185228</td>
<td>0.396165</td>
<td>0.848923</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>2</td>
<td>0</td>
<td>-1.471183</td>
<td>0.183898</td>
<td>1.287285</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>3</td>
<td>0</td>
<td>-2.32937</td>
<td>2.32937</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>-1.38753</td>
<td>-0.320959</td>
<td>0.508592</td>
<td>1.19989</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>0</td>
<td>-2.0367</td>
<td>0.162937</td>
<td>1.87376</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>0</td>
<td>-3.49406</td>
<td>3.49406</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>0</td>
<td>-2.51768</td>
<td>0.148088</td>
<td>2.3696</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>0</td>
<td>-4.49238</td>
<td>4.49238</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>0</td>
<td>-5.39078</td>
<td>5.39078</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>0</td>
<td>0</td>
<td>-0.664115</td>
<td>-0.283203</td>
<td>-0.029261</td>
<td>0.168249</td>
<td>0.332842</td>
<td>0.475486</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>0</td>
<td>1</td>
<td>-0.819637</td>
<td>-0.270955</td>
<td>0.094834</td>
<td>0.379335</td>
<td>0.616423</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>2</td>
<td>0</td>
<td>-1.05913</td>
<td>-0.185228</td>
<td>0.396165</td>
<td>0.848923</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>3</td>
<td>0</td>
<td>-1.471183</td>
<td>0.183898</td>
<td>1.287285</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>4</td>
<td>0</td>
<td>-2.32937</td>
<td>2.32937</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>-1.03102</td>
<td>-0.401685</td>
<td>0.087795</td>
<td>0.495701</td>
<td>0.849207</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>0</td>
<td>-1.38753</td>
<td>-0.320959</td>
<td>0.508592</td>
<td>1.19989</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Estimation of the Scale and location parameters based upon …

<p>| | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1 2</td>
<td>-2.0367</td>
<td>0.162937</td>
<td>1.87376</td>
<td></td>
</tr>
<tr>
<td>1 3</td>
<td>-3.49406</td>
<td>3.49406</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2 0</td>
<td>-1.66639</td>
<td>-0.429484</td>
<td>0.601286</td>
<td>1.49459</td>
</tr>
<tr>
<td>2 1</td>
<td>-2.51768</td>
<td>0.148088</td>
<td>2.3696</td>
<td></td>
</tr>
<tr>
<td>2 2</td>
<td>-4.49238</td>
<td>4.49238</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3 0</td>
<td>-2.95172</td>
<td>0.137322</td>
<td>2.8144</td>
<td></td>
</tr>
<tr>
<td>3 1</td>
<td>-5.39078</td>
<td>5.39078</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4 0</td>
<td>-6.22035</td>
<td>6.22035</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7 0 0</td>
<td>-0.55550</td>
<td>-0.273427</td>
<td>-0.08538</td>
<td>0.600875</td>
</tr>
<tr>
<td>7 0 1</td>
<td>-0.664115</td>
<td>-0.283203</td>
<td>-0.029261</td>
<td>0.168249</td>
</tr>
<tr>
<td>0 2</td>
<td>-0.819637</td>
<td>-0.270955</td>
<td>0.094834</td>
<td>0.379335</td>
</tr>
<tr>
<td>0 3</td>
<td>-1.05913</td>
<td>-0.185228</td>
<td>0.396165</td>
<td>0.848923</td>
</tr>
<tr>
<td>0 4</td>
<td>-1.471183</td>
<td>0.183898</td>
<td>1.287285</td>
<td></td>
</tr>
<tr>
<td>0 5</td>
<td>-2.32937</td>
<td>2.32937</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 0</td>
<td>-0.80961</td>
<td>-0.393654</td>
<td>-0.070136</td>
<td>0.199467</td>
</tr>
<tr>
<td>1 1</td>
<td>-1.03102</td>
<td>-0.401685</td>
<td>0.087795</td>
<td>0.495701</td>
</tr>
<tr>
<td>1 2</td>
<td>-1.38753</td>
<td>-0.320959</td>
<td>0.508592</td>
<td>1.19989</td>
</tr>
<tr>
<td>1 3</td>
<td>-2.0367</td>
<td>0.162937</td>
<td>1.87376</td>
<td></td>
</tr>
<tr>
<td>1 4</td>
<td>-3.49406</td>
<td>3.49406</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2 0</td>
<td>-1.21092</td>
<td>-0.505787</td>
<td>0.081833</td>
<td>0.591087</td>
</tr>
<tr>
<td>2 1</td>
<td>-1.66639</td>
<td>-0.429484</td>
<td>0.601286</td>
<td>1.49459</td>
</tr>
<tr>
<td>2 2</td>
<td>-2.51768</td>
<td>0.148088</td>
<td>2.3696</td>
<td></td>
</tr>
<tr>
<td>2 3</td>
<td>-4.49238</td>
<td>4.49238</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3 0</td>
<td>-1.91874</td>
<td>-0.524225</td>
<td>0.684315</td>
<td>1.75865</td>
</tr>
<tr>
<td>3 1</td>
<td>-2.95172</td>
<td>0.137322</td>
<td>2.8144</td>
<td></td>
</tr>
<tr>
<td>3 2</td>
<td>-5.39078</td>
<td>5.39078</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4 0</td>
<td>-3.35367</td>
<td>0.128907</td>
<td>3.22477</td>
<td></td>
</tr>
<tr>
<td>4 1</td>
<td>-6.22035</td>
<td>6.22035</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5 0</td>
<td>-6.99737</td>
<td>6.99737</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8 0 0</td>
<td>-0.47562</td>
<td>-0.257089</td>
<td>-0.11140</td>
<td>0.001915</td>
</tr>
<tr>
<td>8 0 1</td>
<td>-0.55550</td>
<td>-0.273427</td>
<td>-0.08538</td>
<td>0.608075</td>
</tr>
<tr>
<td>0 2</td>
<td>-0.664115</td>
<td>-0.283203</td>
<td>-0.029261</td>
<td>0.168249</td>
</tr>
<tr>
<td>0 3</td>
<td>-0.819637</td>
<td>-0.270955</td>
<td>0.094834</td>
<td>0.379335</td>
</tr>
<tr>
<td>0 4</td>
<td>-1.05913</td>
<td>-0.185228</td>
<td>0.396165</td>
<td>0.848923</td>
</tr>
<tr>
<td>0 5</td>
<td>-1.471183</td>
<td>0.183898</td>
<td>1.287285</td>
<td></td>
</tr>
<tr>
<td>0 6</td>
<td>-2.32937</td>
<td>2.32937</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 0</td>
<td>-0.66048</td>
<td>-0.364434</td>
<td>-0.134177</td>
<td>0.057706</td>
</tr>
<tr>
<td>1 1</td>
<td>-0.80961</td>
<td>-0.393654</td>
<td>-0.070136</td>
<td>0.199467</td>
</tr>
<tr>
<td>1 2</td>
<td>-1.03102</td>
<td>-0.401685</td>
<td>0.087795</td>
<td>0.495701</td>
</tr>
<tr>
<td>1 3</td>
<td>-1.38753</td>
<td>-0.320959</td>
<td>0.508592</td>
<td>1.19989</td>
</tr>
<tr>
<td>1 4</td>
<td>-2.0367</td>
<td>0.162937</td>
<td>1.87376</td>
<td></td>
</tr>
<tr>
<td>1 5</td>
<td>-3.49406</td>
<td>3.49406</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2 0</td>
<td>-0.934107</td>
<td>-0.480834</td>
<td>-0.103101</td>
<td>0.224255</td>
</tr>
<tr>
<td>2 1</td>
<td>-1.21092</td>
<td>-0.505787</td>
<td>0.081833</td>
<td>0.591087</td>
</tr>
<tr>
<td>2 2</td>
<td>-1.66639</td>
<td>-0.429484</td>
<td>0.601286</td>
<td>1.49459</td>
</tr>
</tbody>
</table>
Riffat, Ahmad and Hirai

<table>
<thead>
<tr>
<th>n</th>
<th>r1</th>
<th>r2</th>
<th>Var(μ)</th>
<th>Var(λ)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>0</td>
<td>0</td>
<td>1.877692</td>
<td>0.698109</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>0</td>
<td>1.199842</td>
<td>0.347687</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>1</td>
<td>1.877692</td>
<td>0.698109</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>0</td>
<td>3.130157</td>
<td>0.819561</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>0</td>
<td>0.947621</td>
<td>0.230502</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>1</td>
<td>1.199842</td>
<td>0.347687</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>0</td>
<td>1.808979</td>
<td>0.398800</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>1</td>
<td>3.130157</td>
<td>0.819561</td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>0</td>
<td>0.810191</td>
<td>0.171994</td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>1</td>
<td>0.947621</td>
<td>0.230502</td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>2</td>
<td>1.199842</td>
<td>0.347687</td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>3</td>
<td>1.877692</td>
<td>0.698109</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>0</td>
<td>1.360867</td>
<td>0.263194</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>1</td>
<td>1.808979</td>
<td>0.398800</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>2</td>
<td>3.130157</td>
<td>0.819561</td>
</tr>
<tr>
<td>6</td>
<td>0</td>
<td>0</td>
<td>4.172572</td>
<td>0.871367</td>
</tr>
<tr>
<td>6</td>
<td>0</td>
<td>1</td>
<td>0.810191</td>
<td>0.171994</td>
</tr>
<tr>
<td>6</td>
<td>0</td>
<td>2</td>
<td>0.947621</td>
<td>0.230502</td>
</tr>
<tr>
<td>6</td>
<td>0</td>
<td>3</td>
<td>1.199842</td>
<td>0.347687</td>
</tr>
<tr>
<td>6</td>
<td>0</td>
<td>4</td>
<td>1.877692</td>
<td>0.698109</td>
</tr>
</tbody>
</table>

Tables VIII

Variances of alternative linear estimate for location and scale based on the Upper record values from the Weibull distribution from complete samples, singly and doubly censored samples.

Each value may be multiplied by $\lambda^2$.
Estimation of the Scale and location parameters based upon …

<table>
<thead>
<tr>
<th>n</th>
<th>r₁</th>
<th>r₂</th>
<th>Var(μ)</th>
<th>Var(λ)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1.128368</td>
<td>0.196147</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>2</td>
<td>1.808979</td>
<td>0.398800</td>
</tr>
<tr>
<td>1</td>
<td>3</td>
<td></td>
<td>3.130157</td>
<td>0.819561</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>1</td>
<td>1.696965</td>
<td>0.279726</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td></td>
<td>2.31055</td>
<td>0.423073</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td></td>
<td>4.172572</td>
<td>0.871367</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>1</td>
<td>2.757687</td>
<td>0.437662</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td></td>
<td>5.101645</td>
<td>0.900458</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>2</td>
<td>5.944975</td>
<td>0.917695</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>3</td>
<td>0.658752</td>
<td>0.113622</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td></td>
<td>0.721585</td>
<td>0.136939</td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>4</td>
<td>1.384589</td>
<td>0.208902</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>3</td>
<td>1.696965</td>
<td>0.279726</td>
</tr>
<tr>
<td>5</td>
<td>2</td>
<td></td>
<td>2.31055</td>
<td>0.423073</td>
</tr>
<tr>
<td>5</td>
<td>3</td>
<td></td>
<td>4.172572</td>
<td>0.871367</td>
</tr>
<tr>
<td>6</td>
<td>0</td>
<td>0</td>
<td>1.997665</td>
<td>0.290319</td>
</tr>
<tr>
<td>6</td>
<td>0</td>
<td>1</td>
<td>2.757687</td>
<td>0.437662</td>
</tr>
<tr>
<td>6</td>
<td>0</td>
<td>2</td>
<td>5.101645</td>
<td>0.900458</td>
</tr>
<tr>
<td>6</td>
<td>0</td>
<td>3</td>
<td>3.175267</td>
<td>0.448269</td>
</tr>
<tr>
<td>7</td>
<td>0</td>
<td>0</td>
<td>6.775801</td>
<td>0.935643</td>
</tr>
<tr>
<td>7</td>
<td>0</td>
<td>1</td>
<td>0.658752</td>
<td>0.113622</td>
</tr>
<tr>
<td>7</td>
<td>0</td>
<td>2</td>
<td>0.721585</td>
<td>0.136939</td>
</tr>
<tr>
<td>7</td>
<td>0</td>
<td>3</td>
<td>0.947621</td>
<td>0.230502</td>
</tr>
<tr>
<td>7</td>
<td>0</td>
<td>4</td>
<td>1.199842</td>
<td>0.347687</td>
</tr>
<tr>
<td>7</td>
<td>0</td>
<td>5</td>
<td>1.877692</td>
<td>0.698109</td>
</tr>
<tr>
<td>7</td>
<td>1</td>
<td>0</td>
<td>0.983138</td>
<td>0.156098</td>
</tr>
<tr>
<td>7</td>
<td>1</td>
<td>1</td>
<td>1.128368</td>
<td>0.196147</td>
</tr>
<tr>
<td>7</td>
<td>1</td>
<td>2</td>
<td>1.360867</td>
<td>0.263194</td>
</tr>
<tr>
<td>7</td>
<td>1</td>
<td>3</td>
<td>1.808979</td>
<td>0.398800</td>
</tr>
<tr>
<td>7</td>
<td>1</td>
<td>4</td>
<td>3.130157</td>
<td>0.819561</td>
</tr>
<tr>
<td>7</td>
<td>2</td>
<td>0</td>
<td>1.384589</td>
<td>0.208902</td>
</tr>
<tr>
<td>7</td>
<td>2</td>
<td>1</td>
<td>1.696965</td>
<td>0.279726</td>
</tr>
<tr>
<td>7</td>
<td>2</td>
<td>2</td>
<td>2.31055</td>
<td>0.423073</td>
</tr>
<tr>
<td>7</td>
<td>2</td>
<td>3</td>
<td>4.172572</td>
<td>0.871367</td>
</tr>
<tr>
<td>7</td>
<td>3</td>
<td>0</td>
<td>1.997665</td>
<td>0.290319</td>
</tr>
<tr>
<td>7</td>
<td>3</td>
<td>1</td>
<td>2.757687</td>
<td>0.437662</td>
</tr>
<tr>
<td>7</td>
<td>3</td>
<td>2</td>
<td>5.101645</td>
<td>0.900458</td>
</tr>
<tr>
<td>7</td>
<td>4</td>
<td>0</td>
<td>3.175267</td>
<td>0.448269</td>
</tr>
<tr>
<td>7</td>
<td>4</td>
<td>1</td>
<td>5.944975</td>
<td>0.917695</td>
</tr>
<tr>
<td>7</td>
<td>5</td>
<td>0</td>
<td>6.775801</td>
<td>0.935643</td>
</tr>
<tr>
<td>7</td>
<td>6</td>
<td>0</td>
<td>0.612183</td>
<td>0.097303</td>
</tr>
<tr>
<td>8</td>
<td>0</td>
<td>0</td>
<td>0.658752</td>
<td>0.113622</td>
</tr>
<tr>
<td>8</td>
<td>0</td>
<td>1</td>
<td>0.721585</td>
<td>0.136939</td>
</tr>
<tr>
<td>8</td>
<td>0</td>
<td>2</td>
<td>0.810191</td>
<td>0.171994</td>
</tr>
<tr>
<td>8</td>
<td>0</td>
<td>3</td>
<td>0.947621</td>
<td>0.230502</td>
</tr>
<tr>
<td>8</td>
<td>0</td>
<td>4</td>
<td>1.199842</td>
<td>0.347687</td>
</tr>
<tr>
<td>8</td>
<td>0</td>
<td>5</td>
<td>1.877692</td>
<td>0.698109</td>
</tr>
<tr>
<td>8</td>
<td>1</td>
<td>0</td>
<td>0.884891</td>
<td>0.130135</td>
</tr>
<tr>
<td>8</td>
<td>1</td>
<td>1</td>
<td>0.983138</td>
<td>0.156098</td>
</tr>
<tr>
<td>8</td>
<td>1</td>
<td>2</td>
<td>1.128368</td>
<td>0.196147</td>
</tr>
<tr>
<td>8</td>
<td>1</td>
<td>3</td>
<td>1.360867</td>
<td>0.263194</td>
</tr>
<tr>
<td>8</td>
<td>1</td>
<td>4</td>
<td>1.808979</td>
<td>0.398800</td>
</tr>
</tbody>
</table>
Efficiency is defined as the ratio of variances of the estimator of BLUEs to variances of the estimator to alternate estimates (Gupta) for all choices of censored data for sample size up to order 8. These efficiency are presented in Table that shows the efficiency in all cases is very high we therefore can replace the Gupta method with Lloyd method which can’t be used for large sample and when the exact variance matrix is not known.

**Table IX**

Shows the efficiency of BLUE to Gupta.
<table>
<thead>
<tr>
<th>n</th>
<th>r1</th>
<th>r2</th>
<th>Var(μ) Lloyd</th>
<th>Var(μ) Gupta</th>
<th>Efficiency (%)</th>
<th>Var(λ) Lloyd</th>
<th>Var(λ) Gupta</th>
<th>Efficiency (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>3</td>
<td>1.877692</td>
<td>1.877692</td>
<td>99.9999</td>
<td>0.698109</td>
<td>0.698109</td>
<td>100.0000</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>1.324241</td>
<td>1.360867</td>
<td>97.31</td>
<td>0.250903</td>
<td>0.263194</td>
<td>95.3301</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1.788482</td>
<td>1.808979</td>
<td>98.867</td>
<td>0.390219</td>
<td>0.398800</td>
<td>97.8485</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>3.130157</td>
<td>3.130157</td>
<td>100.0000</td>
<td>0.819562</td>
<td>0.819561</td>
<td>100.0000</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>2.295287</td>
<td>2.31055</td>
<td>99.339</td>
<td>0.417883</td>
<td>0.423073</td>
<td>98.773</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>4.172593</td>
<td>4.172572</td>
<td>100.0000</td>
<td>0.871365</td>
<td>0.871367</td>
<td>99.9998</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>5.101619</td>
<td>5.101645</td>
<td>99.9995</td>
<td>0.900487</td>
<td>0.900458</td>
<td>99.9999</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>0</td>
<td>0.776567</td>
<td>0.810191</td>
<td>95.8499</td>
<td>0.156726</td>
<td>0.171994</td>
<td>91.122</td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>2</td>
<td>0.916001</td>
<td>0.947621</td>
<td>96.664</td>
<td>0.214143</td>
<td>0.230502</td>
<td>93.62</td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>3</td>
<td>1.173652</td>
<td>1.199842</td>
<td>97.3817</td>
<td>0.332094</td>
<td>0.347687</td>
<td>95.52</td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>4</td>
<td>1.877692</td>
<td>1.877692</td>
<td>99.9999</td>
<td>0.698109</td>
<td>0.698109</td>
<td>100.0000</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>1.083002</td>
<td>1.128368</td>
<td>95.9795</td>
<td>0.182876</td>
<td>0.196147</td>
<td>93.2341</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1.324241</td>
<td>1.360867</td>
<td>97.31</td>
<td>0.250903</td>
<td>0.263194</td>
<td>95.3301</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>1.788482</td>
<td>1.808979</td>
<td>98.867</td>
<td>0.390219</td>
<td>0.398800</td>
<td>97.8485</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>3</td>
<td>3.130157</td>
<td>3.130157</td>
<td>100.0000</td>
<td>0.819562</td>
<td>0.819561</td>
<td>100.0000</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>1.655944</td>
<td>1.696965</td>
<td>97.583</td>
<td>0.269468</td>
<td>0.279726</td>
<td>96.333</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>2.295287</td>
<td>2.31055</td>
<td>99.339</td>
<td>0.417883</td>
<td>0.423073</td>
<td>98.773</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>4.172593</td>
<td>4.172572</td>
<td>100.0000</td>
<td>0.871365</td>
<td>0.871367</td>
<td>99.9998</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>2.745547</td>
<td>2.757687</td>
<td>99.56</td>
<td>0.434147</td>
<td>0.437662</td>
<td>99.1968</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>5.101619</td>
<td>5.101645</td>
<td>99.9995</td>
<td>0.900487</td>
<td>0.900458</td>
<td>99.9999</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>5.944956</td>
<td>5.944975</td>
<td>99.9997</td>
<td>0.917693</td>
<td>0.917695</td>
<td>99.9999</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>0</td>
<td>0.623631</td>
<td>0.658752</td>
<td>94.669</td>
<td>0.100976</td>
<td>0.113622</td>
<td>88.87</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>1</td>
<td>0.686847</td>
<td>0.721585</td>
<td>95.186</td>
<td>0.123002</td>
<td>0.136939</td>
<td>89.82</td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>2</td>
<td>0.776567</td>
<td>0.810191</td>
<td>95.8499</td>
<td>0.156726</td>
<td>0.171994</td>
<td>91.122</td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>3</td>
<td>0.916001</td>
<td>0.947621</td>
<td>96.664</td>
<td>0.214143</td>
<td>0.230502</td>
<td>93.62</td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>4</td>
<td>1.173652</td>
<td>1.199842</td>
<td>97.3817</td>
<td>0.332094</td>
<td>0.347687</td>
<td>95.52</td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>5</td>
<td>1.877692</td>
<td>1.877692</td>
<td>99.9999</td>
<td>0.698109</td>
<td>0.698109</td>
<td>100.0000</td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0.933756</td>
<td>0.983138</td>
<td>94.977</td>
<td>0.143050</td>
<td>0.156098</td>
<td>91.932</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>0</td>
<td>1.083002</td>
<td>1.128368</td>
<td>95.9795</td>
<td>0.182876</td>
<td>0.196147</td>
<td>93.2341</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>0</td>
<td>1.324241</td>
<td>1.360867</td>
<td>97.31</td>
<td>0.250903</td>
<td>0.263194</td>
<td>95.3301</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>0</td>
<td>1.788482</td>
<td>1.808979</td>
<td>98.867</td>
<td>0.390219</td>
<td>0.398800</td>
<td>97.8485</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>4</td>
<td>3.130157</td>
<td>3.130157</td>
<td>100.0000</td>
<td>0.819562</td>
<td>0.819561</td>
<td>100.0000</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>0</td>
<td>1.330726</td>
<td>1.384589</td>
<td>96.1098</td>
<td>0.196908</td>
<td>0.206902</td>
<td>94.258</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>1</td>
<td>1.655944</td>
<td>1.696965</td>
<td>97.583</td>
<td>0.269468</td>
<td>0.279726</td>
<td>96.333</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>2</td>
<td>2.295287</td>
<td>2.31055</td>
<td>99.339</td>
<td>0.417883</td>
<td>0.423073</td>
<td>98.773</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>3</td>
<td>4.172593</td>
<td>4.172572</td>
<td>100.0000</td>
<td>0.871365</td>
<td>0.871367</td>
<td>99.9998</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>0</td>
<td>1.953848</td>
<td>1.997665</td>
<td>97.8066</td>
<td>0.281447</td>
<td>0.290319</td>
<td>96.944</td>
<td></td>
</tr>
<tr>
<td>n</td>
<td>r₁</td>
<td>r₂</td>
<td>Var(μ) Lloyd</td>
<td>Var(μ) Gupta</td>
<td>Efficiency (%)</td>
<td>Var(λ) Lloyd</td>
<td>Var(λ) Gupta</td>
<td>Efficiency (%)</td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>2.745547</td>
<td>2.757687</td>
<td>99.56</td>
<td>0.434147</td>
<td>0.437662</td>
<td>99.1968</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>5.101619</td>
<td>5.101645</td>
<td>99.9995</td>
<td>0.900487</td>
<td>0.900458</td>
<td>99.9999</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>3.166619</td>
<td>3.175267</td>
<td>99.728</td>
<td>0.445984</td>
<td>0.448269</td>
<td>99.490</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>5.944956</td>
<td>5.944975</td>
<td>99.9997</td>
<td>0.917693</td>
<td>0.917695</td>
<td>99.9999</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>6.775784</td>
<td>6.775801</td>
<td>99.99973</td>
<td>0.935642</td>
<td>0.935643</td>
<td>99.9999</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>0</td>
<td>0</td>
<td>0.573132</td>
<td>0.612183</td>
<td>93.621</td>
<td>0.084771</td>
<td>0.097303</td>
<td>87.17</td>
</tr>
<tr>
<td>8</td>
<td>0</td>
<td>1</td>
<td>0.623631</td>
<td>0.658752</td>
<td>94.669</td>
<td>0.100976</td>
<td>0.113622</td>
<td>88.87</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0.686847</td>
<td>0.721585</td>
<td>95.186</td>
<td>0.123002</td>
<td>0.136939</td>
<td>89.82</td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>3</td>
<td>0.776567</td>
<td>0.810191</td>
<td>95.8499</td>
<td>0.156726</td>
<td>0.171994</td>
<td>91.122</td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>4</td>
<td>0.916001</td>
<td>0.947621</td>
<td>96.664</td>
<td>0.214143</td>
<td>0.230502</td>
<td>93.62</td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>5</td>
<td>1.173652</td>
<td>1.199842</td>
<td>97.3817</td>
<td>0.332094</td>
<td>0.347687</td>
<td>95.52</td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>6</td>
<td>1.877692</td>
<td>1.877692</td>
<td>99.9999</td>
<td>0.698109</td>
<td>0.698109</td>
<td>100</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>0.824926</td>
<td>0.884891</td>
<td>93.225</td>
<td>0.115770</td>
<td>0.130135</td>
<td>88.964</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>0.933756</td>
<td>0.983138</td>
<td>94.977</td>
<td>0.143050</td>
<td>0.156098</td>
<td>91.932</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>1.083002</td>
<td>1.128368</td>
<td>95.9795</td>
<td>0.182876</td>
<td>0.196147</td>
<td>93.234</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>3</td>
<td>1.324241</td>
<td>1.360867</td>
<td>97.31</td>
<td>0.250903</td>
<td>0.263194</td>
<td>95.3301</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>4</td>
<td>1.788482</td>
<td>1.808979</td>
<td>98.867</td>
<td>0.390219</td>
<td>0.398800</td>
<td>97.8485</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>5</td>
<td>3.130157</td>
<td>3.130157</td>
<td>100.00001</td>
<td>0.819562</td>
<td>0.819561</td>
<td>100.0000</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>1.120127</td>
<td>1.198084</td>
<td>93.493</td>
<td>0.152114</td>
<td>0.167908</td>
<td>90.594</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>1.330726</td>
<td>1.384589</td>
<td>96.1098</td>
<td>0.196908</td>
<td>0.208902</td>
<td>94.258</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>1.655944</td>
<td>1.696965</td>
<td>97.583</td>
<td>0.296946</td>
<td>0.297926</td>
<td>96.333</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>2.295287</td>
<td>2.31055</td>
<td>99.339</td>
<td>0.417883</td>
<td>0.423073</td>
<td>98.773</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>4</td>
<td>4.172593</td>
<td>4.172572</td>
<td>100.00001</td>
<td>0.871365</td>
<td>0.871367</td>
<td>99.9998</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>1.531166</td>
<td>1.627784</td>
<td>94.0644</td>
<td>0.202719</td>
<td>0.220153</td>
<td>92.0808</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>1.953848</td>
<td>1.997665</td>
<td>97.8066</td>
<td>0.281447</td>
<td>0.293018</td>
<td>96.944</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>2.745547</td>
<td>2.757687</td>
<td>99.56</td>
<td>0.434147</td>
<td>0.437662</td>
<td>99.168</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>5.101619</td>
<td>5.101645</td>
<td>99.9995</td>
<td>0.900487</td>
<td>0.900458</td>
<td>99.9999</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>2.187797</td>
<td>2.306431</td>
<td>94.8564</td>
<td>0.283559</td>
<td>0.303245</td>
<td>93.508</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>3.166619</td>
<td>3.175267</td>
<td>99.728</td>
<td>0.445984</td>
<td>0.448269</td>
<td>99.490</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>2</td>
<td>5.944956</td>
<td>5.944975</td>
<td>99.9997</td>
<td>0.917693</td>
<td>0.917695</td>
<td>99.9999</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>3.472993</td>
<td>3.569157</td>
<td>97.306</td>
<td>0.441927</td>
<td>0.458056</td>
<td>96.479</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>6.775784</td>
<td>6.775801</td>
<td>99.9997</td>
<td>0.935642</td>
<td>0.935643</td>
<td>99.9999</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>0</td>
<td>5.819889</td>
<td>5.819889</td>
<td>99.9999</td>
<td>0.723788</td>
<td>0.723789</td>
<td>99.9999</td>
<td></td>
</tr>
</tbody>
</table>

**ACKNOWLEDGEMENT**

I am very grateful to Professor Abdul-Samad-Hirai former Director and Professor of Statistics Institute of Statistics, University of Punjab under his kind guidance this work has been completed.
REFERENCES


ESTIMATION OF THE PARAMETERS BASED ON LOWER RECORD VALUES FROM THE GUMBEL DISTRIBUTION FOR TYPE II SINGLY AND DOUBLY CENSORED DATA

Saba Aziz1, Azaz Ahmad2 and Abdul Samad Hirai3

1 College of Statistics and Actuarial Science, Punjab University, Lahore, Pakistan. Email: sabaziz41@hotmail.com
2 Govt. Postgraduate College, Okara, Pakistan. Email: azazpu@yahoo.com
3 52 Mamdot Block, Mustafa Town, Lahore, Pakistan

ABSTRACT

The record values play an important role in the solution of various problems, arising in floods, draught predictions, multiple comparison tests and multiple decision procedures. The statistical study of record values started with Chandler (1952). Holland (1994) obtained the minimum variance unbiased estimators (M.U.L.U.E), best linear invariant estimators (B.L.I.E) and maximum likelihood estimators of location and scale parameters of the generalized extreme value distribution using lower record values of complete data.

Ahsanullah (1994) derived the expression for probability density function of lower record values of Gumbel distribution. We use the expected vector and variance covariance matrix of the lower record values to obtain the coefficients of B.L.U.E of the location and scale parameters from singly and doubly censored data with the alternative linear estimate of Gupta method for \( n \geq 8 \).

1. INTRODUCTION

In probability theory and statistics, the Gumbel distribution (named after Emil Julius Gumbel (1891-1966)) is used to model the distribution of the maximum (or the minimum) of a number of samples of various distributions.

The generalized extreme value distribution was described by Jenkinson (1955) considering Gumbel, Frechet and Weibull distributions as special cases of generalized extreme value distribution. Balakrishnan and Chan (1992) obtained the tables of mean, variance and covariance of order statistics for extreme value distribution and also derived best linear unbiased estimators using order statistics of location and scale parameter for extreme value distribution. Ahsanullah and Holland (1994) obtained the Minimum variance linear unbiased estimators (MVLUE), best linear invariant estimators (BLUE) and maximum likelihood estimators of location and scale parameters of the generalized extreme value distribution using lower record values for complete data.

In section 2 of this research article we discuss lower record values of gumbel distribution (maximum). In section 3 coefficients for best linear unbiased estimator of the location and scale parameter of Gumbel distribution based on lower record values from
Type II singly and doubly censored samples for $1 \leq n \leq 8$ is obtained using Lloyds method. In section 4 coefficients of alternative linear estimates (Gupta) is obtained from Type II singly and doubly censored samples. In section 5 relative efficiencies of Lloyds and Gupta are compared.

2. LOWER RECORD VALUES FOR GUMBEL DISTRIBUTION

The extreme value type I distribution is also referred to as the Gumbel distribution. The probability density function of Gumbel distribution is given as:

$$f(x) = \frac{1}{\sigma} e^{-\frac{(x-\mu)}{\sigma}} e^{-e^{-\frac{(x-\mu)}{\sigma}}}, \quad -\infty \leq x \leq \infty$$

Ahsanullah (1994) has derived the following expression for probability density function of lower record values of Gumbel distribution with location parameter $\mu$ and scale parameter $\sigma$.

$$f_r(x) = \frac{1}{(r-1)!} e^{-\frac{r}{\sigma}(x-\mu)} e^{-e^{-\frac{r}{\sigma}(x-\mu)}}, \quad -\infty \leq x \leq \infty$$

$$E(x_{L(r)}) = \mu + \nu_r \sigma$$

$$Var(x_{L(r)}) = \sigma^2 V_{r,r}$$

$$Cov(x_{L(r),x_{L(m)}}) = Var(x_{L(m)}) \text{ for } r < m$$

$$\nu_1^* = \nu = 0.5772 \text{ is Euler's constant.} \quad \nu_j^* = \nu_{j-1}^* - (j-1)^{-1} \quad j \geq 2$$

$$V_{1,1} = \pi^2 / 6 \quad V_{j,j} = V_{j-1,j-1} - (j-1)^{-2} \quad j \geq 2$$

Table I

Means of the lower record values from gumbel distribution for $1 \leq n \leq 8$

<table>
<thead>
<tr>
<th>n</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>0.5772</td>
<td>-0.42278</td>
<td>-0.92278</td>
<td>-1.25611</td>
<td>-1.5061</td>
<td>-1.7061</td>
<td>-1.8728</td>
<td>-2.0156</td>
</tr>
</tbody>
</table>

Table II

Variance covariance of the lower record values from gumbel distribution for $1 \leq n \leq 8$

<table>
<thead>
<tr>
<th>m/r</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.6449</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>0.6449</td>
<td>0.6449</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>0.3949</td>
<td>0.3949</td>
<td>0.3949</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>0.2838</td>
<td>0.2838</td>
<td>0.2838</td>
<td>0.2838</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>0.2213</td>
<td>0.2213</td>
<td>0.2213</td>
<td>0.2213</td>
<td>0.2213</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>0.1813</td>
<td>0.1813</td>
<td>0.1813</td>
<td>0.1813</td>
<td>0.1813</td>
<td>0.1813</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>0.1535</td>
<td>0.1535</td>
<td>0.1535</td>
<td>0.1535</td>
<td>0.1535</td>
<td>0.1535</td>
<td>0.1535</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>0.1331</td>
<td>0.1331</td>
<td>0.1331</td>
<td>0.1331</td>
<td>0.1331</td>
<td>0.1331</td>
<td>0.1331</td>
<td>0.1331</td>
</tr>
</tbody>
</table>

3. ESTIMATION OF $\mu$ AND $\lambda$ BY BLUE (LLOYDS METHOD)

The best linear unbiased estimators $\hat{\mu}$ and $\hat{\sigma}^*$ expressed as the linear function of lower record statistics from Gumbel distribution is given as:
\[ \theta^* = \begin{bmatrix} \mu^* \\ \sigma^* \end{bmatrix} \]

\[ \mu^* = \sum_{i=\tau_1+1}^{n-\tau_2} a_i X_{L(i)}^{(n)} \]

\[ \sigma^* = \sum_{i=1+\tau_1}^{n-\tau_2} b_i X_{L(i)}^{(n)} \]

\[ a = \frac{1}{\Delta} \left( \alpha' \left( V_L^{(n)} \right)^{-1} \alpha' - \left( \left( \left( V_L^{(n)} \right)^{-1} \alpha \right)' \left( V_L^{(n)} \right)^{-1} \right) \right) \]

\[ b = \frac{1}{\Delta} \left( \left( \left( V_L^{(n)} \right)^{-1} \alpha \right)' - \left( \left( \left( V_L^{(n)} \right)^{-1} \alpha \right)' \left( V_L^{(n)} \right)^{-1} \right) \right) \]

\[ \Delta = \left( \left( \left( V_L^{(n)} \right)^{-1} \alpha \right)' \right)^2 \]

where \( \tau_1 \) and \( \tau_2 \) are fixed in advance. \( \tau_1 \) is the missing observation on left and \( \tau_2 \) is the missing observations on right. \( (n-\tau_1, \tau_2) \) are measured observation.

\[ \text{Var}(\theta^*) = \sigma^2 \frac{\alpha'}{\Delta} \begin{bmatrix} \alpha' \left( V_L^{(n)} \right)^{-1} \alpha - \left( \left( V_L^{(n)} \right)^{-1} \right) \alpha' \\ -\left( \left( V_L^{(n)} \right)^{-1} \right) \alpha' \end{bmatrix} \]

\[ X_L = \begin{bmatrix} X_{L(\tau_1+1)}^{(n)} \\ X_{L(\tau_1+2)}^{(n)} \\ \vdots \\ X_{L(n-\tau_1-\tau_2)}^{(n)} \end{bmatrix} \]

and \( \alpha' = \begin{bmatrix} \alpha_{L(\tau_1+1)}^{(n)} \cdots \alpha_{L(n-\tau_1-\tau_2)}^{(n)} \end{bmatrix} \)

\[ V_L^{(n)} = \begin{bmatrix} V_{L(\tau_1+1)(\tau_1+1)}^{(n)} & V_{L(\tau_1+1)(\tau_2+1)}^{(n)} & \cdots & V_{L(\tau_1+1)(n-\tau_1-\tau_2)}^{(n)} \\ V_{(\tau_1+2)(\tau_1+2)}^{(n)} & V_{(\tau_1+2)(\tau_2+2)}^{(n)} & \cdots & V_{(\tau_1+2)(n-\tau_1-\tau_2)}^{(n)} \\ \vdots & \vdots & \ddots & \vdots \\ V_{L(n-\tau_1-\tau_2)(n-\tau_1-\tau_2)}^{(n)} & V_{L(n-\tau_1-\tau_2)(\tau_2+2)}^{(n)} & \cdots & V_{L(n-\tau_1-\tau_2)(n-\tau_1-\tau_2)}^{(n)} \end{bmatrix} \]
Estimation of the parameters based on lower record values...

**Table III**

Showing the coefficients for BLUE of the location parameter based on lower record values of Gumbel distribution from singly and doubly censored samples

<table>
<thead>
<tr>
<th>n</th>
<th>r1</th>
<th>r2</th>
<th>a_{11}</th>
<th>a_{12}</th>
<th>a_{13}</th>
<th>a_{14}</th>
<th>a_{15}</th>
<th>a_{16}</th>
<th>a_{17}</th>
<th>a_{18}</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>0 1</td>
<td>0.42278</td>
<td>0.57722</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>1 0</td>
<td>1.84556</td>
<td>-0.84556</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>0 1</td>
<td>0.46139</td>
<td>0.46140</td>
<td>0.07722</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>0 2</td>
<td>0.42278</td>
<td>0.57722</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>1 0</td>
<td>1.25606</td>
<td>0.62820</td>
<td>-0.88426</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>1 1</td>
<td>1.84556</td>
<td>-0.84556</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>2 0</td>
<td>3.76837</td>
<td>-2.76837</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>0 1</td>
<td>0.41869</td>
<td>0.41869</td>
<td>0.41881</td>
<td>-0.25618</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>0 2</td>
<td>0.46139</td>
<td>0.46140</td>
<td>0.07722</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>0 3</td>
<td>0.42278</td>
<td>0.57722</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>1 0</td>
<td>1.00407</td>
<td>0.50217</td>
<td>0.50182</td>
<td>-1.00805</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>1 1</td>
<td>1.25606</td>
<td>0.62820</td>
<td>-0.88426</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>1 2</td>
<td>1.84556</td>
<td>-0.84556</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>2 0</td>
<td>2.25935</td>
<td>0.75273</td>
<td>-2.01208</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>2 1</td>
<td>3.76837</td>
<td>-2.76837</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>3 0</td>
<td>6.02464</td>
<td>-5.02464</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>0 1</td>
<td>0.37691</td>
<td>0.37691</td>
<td>0.37701</td>
<td>0.37683</td>
<td>-0.50793</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>0 2</td>
<td>0.41869</td>
<td>0.41869</td>
<td>0.41881</td>
<td>-0.25618</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>0 3</td>
<td>0.46139</td>
<td>0.46140</td>
<td>0.07722</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>0 4</td>
<td>0.42278</td>
<td>0.57722</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>1 0</td>
<td>0.85305</td>
<td>0.42664</td>
<td>0.42659</td>
<td>-1.13262</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>1 1</td>
<td>1.00407</td>
<td>0.50217</td>
<td>0.50182</td>
<td>-1.00805</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>1 2</td>
<td>1.25606</td>
<td>0.62820</td>
<td>-0.88426</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>1 3</td>
<td>1.84556</td>
<td>-0.84556</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>2 0</td>
<td>1.70625</td>
<td>0.56846</td>
<td>0.56871</td>
<td>-1.8435</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>2 1</td>
<td>2.25935</td>
<td>0.75273</td>
<td>-2.01208</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>2 2</td>
<td>3.76837</td>
<td>-2.76837</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>3 0</td>
<td>3.41220</td>
<td>0.85322</td>
<td>-3.26542</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>3 1</td>
<td>6.02464</td>
<td>-5.02464</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>4 0</td>
<td>8.5305</td>
<td>-7.5305</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>0 1</td>
<td>0.34122</td>
<td>0.34122</td>
<td>0.34131</td>
<td>0.34107</td>
<td>0.34127</td>
<td>-0.70601</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>0 2</td>
<td>0.37691</td>
<td>0.37691</td>
<td>0.37701</td>
<td>0.37683</td>
<td>-0.50793</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>0 3</td>
<td>0.41869</td>
<td>0.41869</td>
<td>0.41881</td>
<td>-0.25618</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>0 4</td>
<td>0.46139</td>
<td>0.46140</td>
<td>0.07722</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>0 5</td>
<td>0.42278</td>
<td>0.57722</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>1 0</td>
<td>0.74918</td>
<td>0.37469</td>
<td>0.37443</td>
<td>0.37465</td>
<td>0.37324</td>
<td>-1.24619</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>1 1</td>
<td>0.85305</td>
<td>0.42664</td>
<td>0.42634</td>
<td>0.42659</td>
<td>-1.13262</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>1 2</td>
<td>1.00407</td>
<td>0.50217</td>
<td>0.50182</td>
<td>-1.00805</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>1 3</td>
<td>1.25606</td>
<td>0.62820</td>
<td>-0.88426</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>1 4</td>
<td>1.84556</td>
<td>-0.84556</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>2 0</td>
<td>1.40487</td>
<td>0.46805</td>
<td>0.46832</td>
<td>0.46656</td>
<td>-1.8078</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>2 1</td>
<td>1.70625</td>
<td>0.56846</td>
<td>0.56871</td>
<td>-1.8435</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>2 2</td>
<td>2.25935</td>
<td>0.75273</td>
<td>-2.01208</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>2 3</td>
<td>3.76837</td>
<td>-2.76837</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>3 0</td>
<td>2.49737</td>
<td>0.62447</td>
<td>0.62212</td>
<td>-2.74395</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table IV

Showing the coefficients for BLUE of the scale parameter based on lower record values of gumbel distribution from singly and doubly censored samples

<table>
<thead>
<tr>
<th>n</th>
<th>r1</th>
<th>r2</th>
<th>a11</th>
<th>a12</th>
<th>a13</th>
<th>a14</th>
<th>a15</th>
<th>a16</th>
<th>a17</th>
<th>a18</th>
</tr>
</thead>
<tbody>
<tr>
<td>7</td>
<td>3</td>
<td>1</td>
<td>4.1220</td>
<td>8.5322</td>
<td>-3.26542</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>3</td>
<td>2</td>
<td>6.02464</td>
<td>-5.02464</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>4</td>
<td>0</td>
<td>4.68294</td>
<td>0.93322</td>
<td>-4.61615</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>4</td>
<td>1</td>
<td>8.5305</td>
<td>-7.5305</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>5</td>
<td>0</td>
<td>11.2346</td>
<td>-10.2346</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>0</td>
<td>1</td>
<td>0.31215</td>
<td>0.31215</td>
<td>0.31224</td>
<td>0.31202</td>
<td>0.31103</td>
<td>-0.87180</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>0</td>
<td>2</td>
<td>0.34122</td>
<td>0.34122</td>
<td>0.34131</td>
<td>0.34107</td>
<td>0.34127</td>
<td>-0.70601</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>0</td>
<td>3</td>
<td>0.37691</td>
<td>0.37691</td>
<td>0.37701</td>
<td>0.37683</td>
<td>-0.50793</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>0</td>
<td>4</td>
<td>0.41869</td>
<td>0.41869</td>
<td>0.41881</td>
<td>0.41881</td>
<td>-0.25618</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>0</td>
<td>5</td>
<td>0.46139</td>
<td>0.46140</td>
<td>0.07722</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>0</td>
<td>6</td>
<td>0.42278</td>
<td>0.57722</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>1</td>
<td>0</td>
<td>0.67196</td>
<td>0.33607</td>
<td>0.33583</td>
<td>0.33603</td>
<td>0.33477</td>
<td>0.33719</td>
<td>-1.35185</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>1</td>
<td>1</td>
<td>0.74918</td>
<td>0.37469</td>
<td>0.37443</td>
<td>0.37465</td>
<td>0.37324</td>
<td>-1.24619</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>1</td>
<td>2</td>
<td>0.85305</td>
<td>0.42664</td>
<td>0.42634</td>
<td>0.42659</td>
<td>-1.13262</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>1</td>
<td>3</td>
<td>1.00407</td>
<td>0.50217</td>
<td>0.50182</td>
<td>-1.00805</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>1</td>
<td>4</td>
<td>1.25606</td>
<td>0.62820</td>
<td>-0.88426</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>1</td>
<td>5</td>
<td>1.84556</td>
<td>-0.84556</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>2</td>
<td>0</td>
<td>1.20966</td>
<td>0.40301</td>
<td>0.40325</td>
<td>0.40173</td>
<td>-1.82229</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>2</td>
<td>1</td>
<td>1.40487</td>
<td>0.46805</td>
<td>0.46832</td>
<td>0.46665</td>
<td>-1.8078</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>2</td>
<td>2</td>
<td>1.70625</td>
<td>0.56846</td>
<td>0.56871</td>
<td>-1.8435</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>2</td>
<td>3</td>
<td>2.25935</td>
<td>0.75273</td>
<td>-2.01208</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>2</td>
<td>4</td>
<td>3.76837</td>
<td>-2.76837</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>3</td>
<td>0</td>
<td>2.01596</td>
<td>0.50409</td>
<td>0.50220</td>
<td>0.50582</td>
<td>-2.52808</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>3</td>
<td>1</td>
<td>2.49737</td>
<td>0.62447</td>
<td>0.62212</td>
<td>-2.74395</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>3</td>
<td>2</td>
<td>3.41220</td>
<td>0.85322</td>
<td>-3.26542</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>3</td>
<td>3</td>
<td>6.02464</td>
<td>-5.02464</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>4</td>
<td>0</td>
<td>3.36023</td>
<td>0.66963</td>
<td>0.67446</td>
<td>-3.70432</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>4</td>
<td>1</td>
<td>4.68294</td>
<td>0.93322</td>
<td>-4.61615</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>4</td>
<td>2</td>
<td>8.5305</td>
<td>-7.5305</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>5</td>
<td>0</td>
<td>6.04559</td>
<td>1.01183</td>
<td>-6.05742</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>5</td>
<td>1</td>
<td>11.2346</td>
<td>-10.2346</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>6</td>
<td>0</td>
<td>14.1148</td>
<td>-13.1148</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>n</td>
<td>r_1</td>
<td>r_2</td>
<td>b_{11}</td>
<td>b_{12}</td>
<td>b_{13}</td>
<td>b_{14}</td>
<td>b_{15}</td>
<td>b_{16}</td>
<td>b_{17}</td>
<td>b_{18}</td>
</tr>
<tr>
<td>---</td>
<td>-----</td>
<td>-----</td>
<td>-------------</td>
<td>-------------</td>
<td>-------------</td>
<td>-------------</td>
<td>-------------</td>
<td>-------------</td>
<td>-------------</td>
<td>-------------</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>0</td>
<td>0.66667</td>
<td>0.33342</td>
<td>0.33319</td>
<td>-1.33328</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>1</td>
<td>0.99996</td>
<td>0.50012</td>
<td>-1.50008</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>2</td>
<td></td>
<td>2</td>
<td>-2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>2</td>
<td>0</td>
<td>1.50014</td>
<td>0.49978</td>
<td>-1.99992</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>2</td>
<td>1</td>
<td>3.00003</td>
<td>-3.00003</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>3</td>
<td>0</td>
<td></td>
<td></td>
<td>4.00016</td>
<td>-4.00016</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>0</td>
<td>1</td>
<td>0.25031</td>
<td>0.25031</td>
<td>0.25038</td>
<td>0.25021</td>
<td>-1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>0</td>
<td>2</td>
<td>0.33332</td>
<td>0.33332</td>
<td>0.33341</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>0</td>
<td>3</td>
<td>.5</td>
<td>.5</td>
<td>-1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>0</td>
<td>4</td>
<td>1</td>
<td>-1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>0</td>
<td>0.50</td>
<td>0.25007</td>
<td>0.24989</td>
<td>0.25004</td>
<td>-1.24991</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>1</td>
<td>0.66667</td>
<td>0.33342</td>
<td>0.33319</td>
<td>-1.33328</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>2</td>
<td>0.99996</td>
<td>0.500115</td>
<td>-1.50008</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>3</td>
<td>2</td>
<td>-2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>2</td>
<td>0</td>
<td>1.00009</td>
<td>0.33339</td>
<td>0.33339</td>
<td>-1.66667</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>2</td>
<td>1</td>
<td>1.50014</td>
<td>0.49978</td>
<td>-1.99992</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>2</td>
<td>2</td>
<td>3.00003</td>
<td>-3.00003</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>3</td>
<td>0</td>
<td></td>
<td></td>
<td>2</td>
<td>0.50010</td>
<td>-2.50010</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>3</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td>4.00016</td>
<td>-4.00016</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>4</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>5</td>
<td>-5</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>0</td>
<td>1</td>
<td>0.2</td>
<td>0.2</td>
<td>0.20005</td>
<td>0.19991</td>
<td>0.20003</td>
<td>-1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>0</td>
<td>2</td>
<td>0.25031</td>
<td>0.25031</td>
<td>0.25038</td>
<td>0.25021</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>0</td>
<td>3</td>
<td>0.33332</td>
<td>0.33332</td>
<td>0.33341</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>0</td>
<td>4</td>
<td>.5</td>
<td>.5</td>
<td>-1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>0</td>
<td>5</td>
<td>1</td>
<td>-1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>1</td>
<td>0</td>
<td>0.40003</td>
<td>0.20001</td>
<td>0.19993</td>
<td>0.20005</td>
<td>0.19930</td>
<td>-1.19938</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>1</td>
<td>1</td>
<td>0.50</td>
<td>0.25007</td>
<td>0.24989</td>
<td>0.25004</td>
<td>-1.24991</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>1</td>
<td>2</td>
<td>0.66667</td>
<td>0.33342</td>
<td>0.33319</td>
<td>-1.33328</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>1</td>
<td>3</td>
<td>0.99996</td>
<td>0.500115</td>
<td>-1.50008</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>1</td>
<td>4</td>
<td>2</td>
<td>-2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>2</td>
<td>0</td>
<td>0.75014</td>
<td>0.24992</td>
<td>0.25007</td>
<td>0.24913</td>
<td>-1.49925</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>2</td>
<td>1</td>
<td>1.00009</td>
<td>0.33319</td>
<td>0.33339</td>
<td>-1.66667</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>2</td>
<td>2</td>
<td>1.50014</td>
<td>0.49978</td>
<td>-1.99992</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>2</td>
<td>3</td>
<td>3.00003</td>
<td>-3.00003</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>3</td>
<td>0</td>
<td></td>
<td></td>
<td>1.33349</td>
<td>0.33344</td>
<td>0.33219</td>
<td>-1.99912</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>3</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td>2</td>
<td>0.50010</td>
<td>-2.50010</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>3</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td>4.00016</td>
<td>-4.00016</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>4</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2.5005</td>
<td>0.49830</td>
<td>-2.9988</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>4</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>5</td>
</tr>
<tr>
<td>7</td>
<td>5</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>5.9988</td>
</tr>
<tr>
<td>8</td>
<td>0</td>
<td>1</td>
<td>0.16668</td>
<td>0.16668</td>
<td>0.16672</td>
<td>0.16670</td>
<td>0.16608</td>
<td>-0.99947</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>0</td>
<td>2</td>
<td>0.2</td>
<td>0.2</td>
<td>0.20005</td>
<td>0.19991</td>
<td>0.20003</td>
<td>-1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>0</td>
<td>3</td>
<td>0.25031</td>
<td>0.25031</td>
<td>0.25038</td>
<td>0.25021</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>0</td>
<td>4</td>
<td>0.33332</td>
<td>0.33332</td>
<td>0.33341</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>0</td>
<td>5</td>
<td>.5</td>
<td>.5</td>
<td>-1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>0</td>
<td>6</td>
<td>1</td>
<td>-1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>1</td>
<td>0</td>
<td>0.33338</td>
<td>0.16673</td>
<td>0.16662</td>
<td>0.16672</td>
<td>0.16609</td>
<td>0.16729</td>
<td>-1.16682</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>1</td>
<td>1</td>
<td>0.40003</td>
<td>0.20001</td>
<td>0.19993</td>
<td>0.20005</td>
<td>0.19930</td>
<td>-1.19938</td>
<td></td>
<td></td>
</tr>
<tr>
<td>n</td>
<td>r₁</td>
<td>r₂</td>
<td>b₁₁</td>
<td>b₁₂</td>
<td>b₁₃</td>
<td>b₁₄</td>
<td>b₁₅</td>
<td>b₁₆</td>
<td>b₁₇</td>
<td>b₁₈</td>
</tr>
<tr>
<td>---</td>
<td>----</td>
<td>----</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
</tr>
<tr>
<td>8</td>
<td>1 2</td>
<td>0.50</td>
<td>0.25007</td>
<td>0.24989</td>
<td>0.25004</td>
<td>-1.24991</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>1 3</td>
<td>0.666667</td>
<td>0.333423</td>
<td>0.33319</td>
<td>-1.33325</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>1 4</td>
<td>0.99996</td>
<td>0.500115</td>
<td>-1.50008</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>1 5</td>
<td>2</td>
<td>-2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>2 0</td>
<td>0.60015</td>
<td>0.19995</td>
<td>0.20006</td>
<td>0.19931</td>
<td>0.20075</td>
<td>-1.40022</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>2 1</td>
<td>0.75014</td>
<td>0.24992</td>
<td>0.25007</td>
<td>0.24913</td>
<td>-1.49925</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>2 2</td>
<td>1.00009</td>
<td>0.33319</td>
<td>0.33339</td>
<td>-1.66667</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>2 3</td>
<td>1.50014</td>
<td>0.49978</td>
<td>-1.99992</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>2 4</td>
<td>3.00003</td>
<td>-3.00003</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>3 0</td>
<td>1.00018</td>
<td>0.25010</td>
<td>0.24916</td>
<td>0.25095</td>
<td>-1.75038</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>3 1</td>
<td>1.33349</td>
<td>0.33344</td>
<td>0.33219</td>
<td>-1.99912</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>3 2</td>
<td>2</td>
<td>0.50010</td>
<td>-2.50010</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>3 3</td>
<td>4.00016</td>
<td>-4.00016</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>4 0</td>
<td>1.66711</td>
<td>0.33222</td>
<td>0.33462</td>
<td>-2.33396</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>4 1</td>
<td>2.5005</td>
<td>0.49830</td>
<td>-2.9988</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>4 2</td>
<td>5</td>
<td>-5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>5 0</td>
<td>2.9994</td>
<td>0.50200</td>
<td>-3.5014</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>5 1</td>
<td>5.9988</td>
<td>-5.9988</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>6 0</td>
<td>7.0028</td>
<td>-7.0028</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

4. AN ALTERNATIVE LINEAR ESTIMATE (GUPTA 1952)

Lloyd’s procedure requires full knowledge of the expectations and the variance covariance matrix of the record values. The covariance be specially may be different to determine. Gupta (1952) has proposed a very simple method applicable when only the expectations are known. The coefficients of these linear estimates are obtained by assuming the variance matrix to be a unit matrix. Let the linear estimates be

\[
\mu^*_1 = \sum_{i=1+τ_1}^{n-τ_2} a_i X_{L(i)}
\]

\[
σ^*_1 = \sum_{i=1+τ_1}^{n-τ_2} b_i X_{L(i)}
\]

\[
a_i = \frac{1}{n-τ_1 - r_2} - \frac{\bar{μ}_k (μ_{L(i)} - \bar{μ}_{L(k)})}{n-τ_2 \sum_{i=1+τ_1}^{n-τ_2} (μ_{L(i)} - \bar{μ}_{L(k)})^2}
\]

\[
b_i = \frac{(μ_{L(i)} - \bar{μ}_{L(k)})}{\sum_{i=1+τ_1}^{n-τ_2} (μ_{L(i)} - \bar{μ}_{L(k)})^2}
\]

\[
\bar{μ}_{L(k)} = \frac{1}{n-τ_1 - r_2} \sum_{i=1+τ_1}^{n-τ_2} μ_{L(i)}
\]

In matrix notation the variance of the estimates can also be written as
\[ \text{Var}(\mu^*_1) = \sigma^2_a V_a \]
\[ \text{Var}(\sigma^*_1) = \sigma^2_b V_b \]

Table V

Showing the coefficients for alternative linear estimate of location parameters based on lower Records values of gumbel distribution from singly and doubly censored samples

<table>
<thead>
<tr>
<th>n</th>
<th>r_1</th>
<th>r_2</th>
<th>a_1</th>
<th>a_2</th>
<th>a_3</th>
<th>a_4</th>
<th>a_5</th>
<th>a_6</th>
<th>a_7</th>
<th>a_8</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>0</td>
<td>1</td>
<td>0.42278</td>
<td>0.57722</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>0</td>
<td>1.84556</td>
<td>-0.84556</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>1</td>
<td>0.51627</td>
<td>0.29675</td>
<td>0.18698</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>2</td>
<td>0.42278</td>
<td>0.57722</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>0</td>
<td>1.42878</td>
<td>0.19640</td>
<td>-0.62518</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>1</td>
<td>1.84556</td>
<td>-0.84556</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>2</td>
<td>0</td>
<td>3.76837</td>
<td>-2.76837</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>1</td>
<td>0.53606</td>
<td>0.27201</td>
<td>0.13998</td>
<td>0.05196</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>2</td>
<td>0.51627</td>
<td>0.29675</td>
<td>0.18698</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>3</td>
<td>0.42278</td>
<td>0.57722</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>0</td>
<td>1.19296</td>
<td>0.41257</td>
<td>-0.10768</td>
<td>-0.49786</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>1</td>
<td>1.42878</td>
<td>0.19640</td>
<td>-0.62518</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>2</td>
<td>1.84556</td>
<td>-0.84556</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>2</td>
<td>0</td>
<td>2.52447</td>
<td>0.13412</td>
<td>-1.65859</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>2</td>
<td>1</td>
<td>3.76837</td>
<td>-2.76837</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>3</td>
<td>0</td>
<td>6.02464</td>
<td>-5.02464</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>0</td>
<td>1</td>
<td>0.53356</td>
<td>0.27364</td>
<td>0.14368</td>
<td>0.05704</td>
<td>-0.00793</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>0</td>
<td>2</td>
<td>0.53606</td>
<td>0.27201</td>
<td>0.13998</td>
<td>0.05196</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>0</td>
<td>3</td>
<td>0.51627</td>
<td>0.29675</td>
<td>0.18698</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>0</td>
<td>4</td>
<td>0.42278</td>
<td>0.57722</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>0</td>
<td>1.03784</td>
<td>0.47173</td>
<td>0.09432</td>
<td>-0.18872</td>
<td>-0.41517</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>1</td>
<td>1.19296</td>
<td>0.41257</td>
<td>-0.10768</td>
<td>-0.49786</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>2</td>
<td>1.42878</td>
<td>0.19640</td>
<td>-0.62518</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>3</td>
<td>1.84556</td>
<td>-0.84556</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>2</td>
<td>0</td>
<td>1.92245</td>
<td>0.61072</td>
<td>-0.37306</td>
<td>-1.16011</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>2</td>
<td>1</td>
<td>2.52447</td>
<td>0.13412</td>
<td>-1.65859</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>2</td>
<td>2</td>
<td>3.76837</td>
<td>-2.76837</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>3</td>
<td>0</td>
<td>3.75177</td>
<td>0.08920</td>
<td>-2.84097</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>3</td>
<td>1</td>
<td>6.02464</td>
<td>-5.02464</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>4</td>
<td>0</td>
<td>8.53050</td>
<td>-7.53050</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>0</td>
<td>1</td>
<td>0.52316</td>
<td>0.2773</td>
<td>0.15437</td>
<td>0.07242</td>
<td>0.01096</td>
<td>-0.03821</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>0</td>
<td>2</td>
<td>0.53356</td>
<td>0.27364</td>
<td>0.14368</td>
<td>0.05704</td>
<td>-0.00793</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>0</td>
<td>3</td>
<td>0.53606</td>
<td>0.27201</td>
<td>0.13998</td>
<td>0.05196</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>0</td>
<td>4</td>
<td>0.51627</td>
<td>0.29675</td>
<td>0.18698</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>0</td>
<td>5</td>
<td>0.42278</td>
<td>0.57722</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>1</td>
<td>0</td>
<td>0.92655</td>
<td>0.48390</td>
<td>0.18880</td>
<td>-0.03252</td>
<td>-0.20958</td>
<td>-0.35716</td>
<td></td>
<td></td>
</tr>
<tr>
<td>n</td>
<td>r₁</td>
<td>r₂</td>
<td>a₁</td>
<td>a₂</td>
<td>a₃</td>
<td>a₄</td>
<td>a₅</td>
<td>a₆</td>
<td>a₇</td>
<td>a₈</td>
</tr>
<tr>
<td>---</td>
<td>----</td>
<td>----</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
</tr>
<tr>
<td>7</td>
<td>1</td>
<td>1</td>
<td>1.0378</td>
<td>0.47173</td>
<td>0.09432</td>
<td>-0.18872</td>
<td>-0.41517</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>1</td>
<td>2</td>
<td>1.19296</td>
<td>0.41257</td>
<td>-0.10768</td>
<td>-0.49786</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>1</td>
<td>3</td>
<td>1.42878</td>
<td>0.19640</td>
<td>-0.62518</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>1</td>
<td>4</td>
<td>1.84556</td>
<td>-0.84556</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>2</td>
<td>0</td>
<td>1.56759</td>
<td>0.707478</td>
<td>0.06241</td>
<td>-0.45367</td>
<td>-0.88382</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>2</td>
<td>1</td>
<td>1.92245</td>
<td>0.61072</td>
<td>-0.37306</td>
<td>-1.16011</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>2</td>
<td>2</td>
<td>2.52447</td>
<td>0.13412</td>
<td>-1.65859</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>2</td>
<td>3</td>
<td>3.76837</td>
<td>-2.76837</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>3</td>
<td>0</td>
<td>2.71273</td>
<td>0.84238</td>
<td>-0.65396</td>
<td>-1.90115</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>3</td>
<td>1</td>
<td>3.75177</td>
<td>0.08920</td>
<td>-2.84097</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>3</td>
<td>2</td>
<td>6.02464</td>
<td>-5.02464</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>4</td>
<td>0</td>
<td>5.08250</td>
<td>0.05427</td>
<td>-4.13677</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>4</td>
<td>1</td>
<td>8.53050</td>
<td>-7.53050</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>5</td>
<td>0</td>
<td>11.2346</td>
<td>-10.2346</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>0</td>
<td>1</td>
<td>0.50993</td>
<td>0.27948</td>
<td>0.16426</td>
<td>0.08744</td>
<td>0.02983</td>
<td>-0.01626</td>
<td>-0.05468</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>0</td>
<td>2</td>
<td>0.52316</td>
<td>0.2773</td>
<td>0.15437</td>
<td>0.07242</td>
<td>0.01096</td>
<td>-0.03821</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>0</td>
<td>3</td>
<td>0.53356</td>
<td>0.27364</td>
<td>0.14368</td>
<td>0.05704</td>
<td>-0.00793</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>0</td>
<td>4</td>
<td>0.53606</td>
<td>0.27201</td>
<td>0.13998</td>
<td>0.05196</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>0</td>
<td>5</td>
<td>0.51627</td>
<td>0.29675</td>
<td>0.18698</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>0</td>
<td>6</td>
<td>0.42278</td>
<td>0.57722</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>1</td>
<td>0</td>
<td>0.84200</td>
<td>0.47912</td>
<td>0.23717</td>
<td>0.05571</td>
<td>-0.08946</td>
<td>-0.21047</td>
<td>-0.31412</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>1</td>
<td>1</td>
<td>0.92655</td>
<td>0.48390</td>
<td>0.18880</td>
<td>-0.03252</td>
<td>-0.20958</td>
<td>-0.35716</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>1</td>
<td>2</td>
<td>1.03784</td>
<td>0.47173</td>
<td>0.09432</td>
<td>-0.18872</td>
<td>-0.41517</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>1</td>
<td>3</td>
<td>1.19296</td>
<td>0.41257</td>
<td>-0.10768</td>
<td>-0.49785</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>1</td>
<td>4</td>
<td>1.42878</td>
<td>0.19640</td>
<td>-0.62518</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>1</td>
<td>5</td>
<td>1.84556</td>
<td>-0.84556</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>2</td>
<td>0</td>
<td>1.33327</td>
<td>0.70989</td>
<td>0.24237</td>
<td>-0.13166</td>
<td>-0.44341</td>
<td>-0.71047</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>2</td>
<td>1</td>
<td>1.56760</td>
<td>0.70748</td>
<td>0.06241</td>
<td>-0.45367</td>
<td>-0.88382</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>2</td>
<td>2</td>
<td>1.92245</td>
<td>0.61072</td>
<td>-0.37306</td>
<td>-1.16011</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>2</td>
<td>3</td>
<td>2.52447</td>
<td>0.13412</td>
<td>-1.65859</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>2</td>
<td>4</td>
<td>3.76837</td>
<td>-2.76837</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>3</td>
<td>0</td>
<td>2.12760</td>
<td>0.96709</td>
<td>0.03865</td>
<td>-0.73522</td>
<td>-1.39813</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>3</td>
<td>1</td>
<td>2.71273</td>
<td>0.84238</td>
<td>-0.65396</td>
<td>-1.90115</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>3</td>
<td>2</td>
<td>3.75177</td>
<td>0.08920</td>
<td>-2.84097</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>3</td>
<td>3</td>
<td>6.02464</td>
<td>-5.02464</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>4</td>
<td>0</td>
<td>3.55505</td>
<td>1.09822</td>
<td>-0.94956</td>
<td>-2.70372</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>4</td>
<td>1</td>
<td>5.0825</td>
<td>0.05427</td>
<td>-4.13677</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>4</td>
<td>2</td>
<td>8.53050</td>
<td>-7.53050</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>5</td>
<td>0</td>
<td>6.50148</td>
<td>0.02376</td>
<td>-5.52524</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>5</td>
<td>1</td>
<td>11.2346</td>
<td>-10.2346</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>6</td>
<td>0</td>
<td>14.1148</td>
<td>-13.1148</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### Table VI

Showing the coefficients for alternative linear estimate of scale parameter based on lower record values of gumbel distribution from singly and doubly censored samples

<table>
<thead>
<tr>
<th>n</th>
<th>r₁</th>
<th>r₂</th>
<th>b₁</th>
<th>b₂</th>
<th>b₃</th>
<th>b₄</th>
<th>b₅</th>
<th>b₆</th>
<th>b₇</th>
<th>b₈</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>0</td>
<td>1</td>
<td>-1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>1</td>
<td>0.71429</td>
<td>-0.14286</td>
<td>-0.57143</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>2</td>
<td>1.26316</td>
<td>-0.15790</td>
<td>-1.10526</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>2</td>
<td>0</td>
<td></td>
<td>3.00003</td>
<td>-3.00003</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>1</td>
<td>0.56522</td>
<td>0.04348</td>
<td>-0.21739</td>
<td>-0.39130</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>2</td>
<td>0.71429</td>
<td>-0.14286</td>
<td>-0.57143</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>3</td>
<td>1</td>
<td>-1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>0</td>
<td>0.91822</td>
<td>0.15831</td>
<td>-0.34829</td>
<td>-0.72823</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>1</td>
<td>1.26316</td>
<td>-0.15790</td>
<td>-1.10526</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>2</td>
<td></td>
<td></td>
<td>2</td>
<td>-2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>2</td>
<td>0</td>
<td></td>
<td>1.78383</td>
<td>-0.16218</td>
<td>-1.62165</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>2</td>
<td>1</td>
<td></td>
<td>3.00003</td>
<td>-3.00003</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>3</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>4.00016</td>
<td>-4.00016</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>0</td>
<td>1</td>
<td>0.47240</td>
<td>0.10429</td>
<td>-0.07976</td>
<td>-0.20246</td>
<td>-0.29448</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>0</td>
<td>2</td>
<td>0.56522</td>
<td>0.04348</td>
<td>-0.21739</td>
<td>-0.39130</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>0</td>
<td>3</td>
<td>0.71429</td>
<td>-0.14286</td>
<td>-0.57143</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>0</td>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td>-1</td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>0</td>
<td>0.72056</td>
<td>0.23369</td>
<td>-0.09088</td>
<td>-0.33431</td>
<td>-0.52905</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>1</td>
<td>0.91822</td>
<td>0.15831</td>
<td>-0.34829</td>
<td>-0.72823</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>2</td>
<td>1.26316</td>
<td>-0.15790</td>
<td>-1.10526</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>3</td>
<td>2</td>
<td>-2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>2</td>
<td>0</td>
<td></td>
<td>1.24090</td>
<td>0.26764</td>
<td>-0.46229</td>
<td>-1.04625</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>2</td>
<td>1</td>
<td></td>
<td>1.78383</td>
<td>-0.16218</td>
<td>-1.62165</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>2</td>
<td>2</td>
<td></td>
<td>3.00003</td>
<td>-3.00003</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>3</td>
<td>0</td>
<td></td>
<td></td>
<td>2.29512</td>
<td>-0.16391</td>
<td>-2.13121</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>3</td>
<td>1</td>
<td></td>
<td></td>
<td>4.00016</td>
<td>-4.00016</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>4</td>
<td>0</td>
<td></td>
<td></td>
<td>5</td>
<td>-5</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>0</td>
<td>1</td>
<td>0.40845</td>
<td>0.12676</td>
<td>-0.01409</td>
<td>-0.10798</td>
<td>-0.17840</td>
<td>-0.23474</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>0</td>
<td>2</td>
<td>0.47240</td>
<td>0.10429</td>
<td>-0.07976</td>
<td>-0.20246</td>
<td>-0.29448</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>0</td>
<td>3</td>
<td>0.56522</td>
<td>0.04348</td>
<td>-0.21739</td>
<td>-0.39130</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>0</td>
<td>4</td>
<td>0.71429</td>
<td>-0.14286</td>
<td>-0.57143</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>0</td>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td>-1</td>
</tr>
<tr>
<td>7</td>
<td>1</td>
<td>0</td>
<td>0.59314</td>
<td>0.24762</td>
<td>0.01728</td>
<td>-0.15548</td>
<td>-0.29368</td>
<td>-0.40888</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>1</td>
<td>1</td>
<td>0.72056</td>
<td>0.23369</td>
<td>-0.09088</td>
<td>-0.33431</td>
<td>-0.52905</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>1</td>
<td>2</td>
<td>0.91822</td>
<td>0.15831</td>
<td>-0.34829</td>
<td>-0.72823</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>1</td>
<td>3</td>
<td>1.26316</td>
<td>-0.15790</td>
<td>-1.10526</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>n</td>
<td>r1</td>
<td>r2</td>
<td>b1</td>
<td>b2</td>
<td>b3</td>
<td>b4</td>
<td>b5</td>
<td>b6</td>
<td>b7</td>
<td>b8</td>
</tr>
<tr>
<td>---</td>
<td>----</td>
<td>----</td>
<td>------</td>
<td>------</td>
<td>--------</td>
<td>--------</td>
<td>--------</td>
<td>--------</td>
<td>--------</td>
<td>--------</td>
</tr>
<tr>
<td>7</td>
<td>1</td>
<td>4</td>
<td>2</td>
<td>-2</td>
<td>-0.09471</td>
<td>-0.44994</td>
<td>-0.74603</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>2</td>
<td>1</td>
<td>1.24090</td>
<td>0.26764</td>
<td>-0.46229</td>
<td>-1.04625</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>2</td>
<td>2</td>
<td>1.78383</td>
<td>-0.16218</td>
<td>-1.62165</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>2</td>
<td>3</td>
<td>3.00003</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-3.00003</td>
</tr>
<tr>
<td>7</td>
<td>3</td>
<td>0</td>
<td>1.55350</td>
<td>0.37368</td>
<td>-0.57022</td>
<td>-1.35696</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>3</td>
<td>1</td>
<td>2.29512</td>
<td>-0.16391</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-2.13121</td>
</tr>
<tr>
<td>7</td>
<td>3</td>
<td>2</td>
<td>4.00016</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-4.00016</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>4</td>
<td>0</td>
<td>2.80187</td>
<td>-0.16464</td>
<td>-2.63723</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>4</td>
<td>1</td>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-5</td>
</tr>
<tr>
<td>7</td>
<td>5</td>
<td>0</td>
<td>5.99880</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-5.99880</td>
</tr>
<tr>
<td>8</td>
<td>0</td>
<td>1</td>
<td>0.36143</td>
<td>0.13452</td>
<td>0.02107</td>
<td>-0.11129</td>
<td>-0.15667</td>
<td>-0.19450</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>0</td>
<td>2</td>
<td>0.40845</td>
<td>0.12676</td>
<td>-0.01409</td>
<td>-0.10798</td>
<td>-0.17840</td>
<td>-0.23474</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>0</td>
<td>3</td>
<td>0.47240</td>
<td>0.10429</td>
<td>-0.07976</td>
<td>-0.20246</td>
<td>-0.29448</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>0</td>
<td>4</td>
<td>0.56522</td>
<td>0.04348</td>
<td>-0.21739</td>
<td>-0.39130</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>0</td>
<td>5</td>
<td>0.71429</td>
<td>-0.14286</td>
<td>-0.57143</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>0</td>
<td>6</td>
<td>1</td>
<td>-1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>1</td>
<td>0</td>
<td>0.50446</td>
<td>0.24261</td>
<td>0.06804</td>
<td>-0.06288</td>
<td>-0.16762</td>
<td>-0.25492</td>
<td>-0.32970</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>1</td>
<td>1</td>
<td>0.59314</td>
<td>0.24762</td>
<td>0.01728</td>
<td>-0.15548</td>
<td>-0.29368</td>
<td>-0.40888</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>1</td>
<td>2</td>
<td>0.72056</td>
<td>0.23369</td>
<td>-0.09088</td>
<td>-0.33431</td>
<td>-0.52905</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>1</td>
<td>3</td>
<td>0.91822</td>
<td>0.15831</td>
<td>-0.34829</td>
<td>-0.72823</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>1</td>
<td>4</td>
<td>1.26316</td>
<td></td>
<td>-0.15790</td>
<td>-1.10526</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>1</td>
<td>5</td>
<td>2</td>
<td>-2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>2</td>
<td>0</td>
<td>0.75431</td>
<td>0.35124</td>
<td>0.04895</td>
<td>-0.19289</td>
<td>-0.39447</td>
<td>-0.56714</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>2</td>
<td>1</td>
<td>0.94137</td>
<td>0.34932</td>
<td>-0.09471</td>
<td>-0.44994</td>
<td>-0.74603</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>2</td>
<td>2</td>
<td>1.24090</td>
<td>0.26764</td>
<td>-0.46229</td>
<td>-1.04625</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>2</td>
<td>3</td>
<td>1.78383</td>
<td>-0.16218</td>
<td>-1.62165</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>2</td>
<td>4</td>
<td>3.00003</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-3.00003</td>
</tr>
<tr>
<td>8</td>
<td>3</td>
<td>0</td>
<td>1.15333</td>
<td>0.45897</td>
<td>-0.09654</td>
<td>-0.55956</td>
<td>-0.95619</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>3</td>
<td>1</td>
<td>1.55350</td>
<td>0.37368</td>
<td>-0.57022</td>
<td>-1.35696</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>3</td>
<td>2</td>
<td>2.29512</td>
<td>-0.16391</td>
<td>-2.13121</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>3</td>
<td>3</td>
<td>4.00016</td>
<td></td>
<td>-4.00016</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>4</td>
<td>0</td>
<td>1.86184</td>
<td>0.47783</td>
<td>-0.67574</td>
<td>-1.66393</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>4</td>
<td>1</td>
<td>2.80187</td>
<td>-0.164641</td>
<td>-2.63723</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>4</td>
<td>2</td>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-5</td>
</tr>
<tr>
<td>8</td>
<td>5</td>
<td>0</td>
<td>3.30761</td>
<td></td>
<td>-0.16601</td>
<td>-3.14161</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>5</td>
<td>1</td>
<td>5.99880</td>
<td></td>
<td>-5.99880</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>6</td>
<td>0</td>
<td>7.00280</td>
<td></td>
<td>-7.00280</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### 5. RELATIVE EFFICIENCY

Efficiency is defined as the ratio of variances of the estimator of BLUEs to variances of the estimator to alternate estimates (Gupta) for all choices of censored data for sample size up to order 8.
### Table VII

<table>
<thead>
<tr>
<th>n</th>
<th>r₁</th>
<th>r₂</th>
<th>Var(μ*) BLUE</th>
<th>Var(μ*) Gupta</th>
<th>Efficiency %</th>
<th>Var(σ*) BLUE</th>
<th>Var(σ*) Gupta</th>
<th>Efficiency %</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>0</td>
<td>1</td>
<td>0.82364 0.82364</td>
<td>100.00000 1</td>
<td>1 1</td>
<td>1 1</td>
<td>1 1</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>0</td>
<td>1.24642 1.24642</td>
<td>100.00000 1</td>
<td>1 1</td>
<td>1 1</td>
<td>1 1</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>1</td>
<td>0.82066 0.82669</td>
<td>99.27059 0.5</td>
<td>0.59184 84.48229</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>2</td>
<td>0.82364 0.82364</td>
<td>100.00000 1</td>
<td>1 1</td>
<td>1 1</td>
<td>1 1</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>0</td>
<td>1.07267 1.08759</td>
<td>98.62816 0.49998 0.534617 93.52116</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>1</td>
<td>1.24642 1.24642</td>
<td>100.00000 1</td>
<td>1 1</td>
<td>1 1</td>
<td>1 1</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>2</td>
<td>0</td>
<td>1.86149 1.86149</td>
<td>100.00000 0.99992 0.99992 100.00000</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>1</td>
<td>0.80972 0.83426</td>
<td>97.05847 0.33332 0.42911 77.67705</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>2</td>
<td>0.82066 0.82669</td>
<td>99.27059 0.5</td>
<td>0.59184 84.48229</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>3</td>
<td>0.82364 0.82364</td>
<td>100.00000 1</td>
<td>1 1</td>
<td>1 1</td>
<td>1 1</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>0</td>
<td>0.97741 1.0037</td>
<td>97.38069 0.33333 0.372682 89.44086</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>1</td>
<td>1.07267 1.08759</td>
<td>98.62816 0.49998 0.534617 93.52116</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>2</td>
<td>1.24642 1.24642</td>
<td>100.00000 1</td>
<td>1 1</td>
<td>1 1</td>
<td>1 1</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>2</td>
<td>0</td>
<td>1.35547 1.37109</td>
<td>98.86076 0.5</td>
<td>0.517885 96.54653</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>2</td>
<td>1</td>
<td>1.86149 1.86149</td>
<td>100.00000 0.99992 0.99992 100.00000</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>3</td>
<td>0</td>
<td>2.48982 2.48982</td>
<td>100.00000 1.00008 1 100.00800</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>0</td>
<td>1</td>
<td>0.78937 0.83284</td>
<td>94.78051 0.25031 0.33916 73.80292</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>0</td>
<td>2</td>
<td>0.80972 0.83426</td>
<td>97.05847 0.33332 0.42911 77.67705</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>0</td>
<td>3</td>
<td>0.82066 0.82669</td>
<td>99.27059 0.5</td>
<td>0.59184 84.48229</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>0</td>
<td>4</td>
<td>0.82364 0.82364</td>
<td>100.00000 1</td>
<td>1 1</td>
<td>1 1</td>
<td>1 1</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>0</td>
<td>0.90899 0.944644</td>
<td>96.22567 0.24991 0.28875 86.54892</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>1</td>
<td>0.97741 1.0037</td>
<td>97.38069 0.33333 0.372682 89.44086</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>2</td>
<td>1.07267 1.08759</td>
<td>98.62816 0.49998 0.534617 93.52116</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>3</td>
<td>1.24642 1.24642</td>
<td>100.00000 1</td>
<td>1 1</td>
<td>1 1</td>
<td>1 1</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>2</td>
<td>0</td>
<td>1.15156 1.17961</td>
<td>97.62210 0.33333 0.357092 93.34569</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>2</td>
<td>1</td>
<td>1.35547 1.37109</td>
<td>98.86076 0.5</td>
<td>0.517885 96.54653</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>2</td>
<td>2</td>
<td>1.86149 1.86149</td>
<td>100.00000 0.99992 0.99992 100.00000</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>3</td>
<td>0</td>
<td>1.63675 1.65116</td>
<td>99.12728 0.50002 0.510907 97.86908</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>3</td>
<td>1</td>
<td>2.48982 2.48982</td>
<td>100.00000 1.00008 1 100.00800</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>4</td>
<td>0</td>
<td>3.09208 3.09208</td>
<td>100.00000 1</td>
<td>1 1</td>
<td>1 1</td>
<td>1 1</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>0</td>
<td>1</td>
<td>0.76346 0.82553</td>
<td>92.48119 0.2  0.28149 71.05048</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>0</td>
<td>2</td>
<td>0.78937 0.83284</td>
<td>94.78051 0.25031 0.33916 73.80292</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>0</td>
<td>3</td>
<td>0.80972 0.83426</td>
<td>97.05847 0.33332 0.42911 77.67705</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>0</td>
<td>4</td>
<td>0.82066 0.82669</td>
<td>99.27059 0.5</td>
<td>0.59184 84.48229</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>0</td>
<td>5</td>
<td>0.82364 0.82364</td>
<td>100.00000 1</td>
<td>1 1</td>
<td>1 1</td>
<td>1 1</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>1</td>
<td>0</td>
<td>0.85503 0.898382</td>
<td>95.17444 0.20002 0.236896 84.43368</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>1</td>
<td>1</td>
<td>0.90899 0.944644</td>
<td>96.22567 0.24991 0.28875 86.54892</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>1</td>
<td>2</td>
<td>0.97741 1.0037</td>
<td>97.38069 0.33333 0.372682 89.44086</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>1</td>
<td>3</td>
<td>1.07267 1.08759</td>
<td>98.62816 0.49998 0.534617 93.52116</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>1</td>
<td>4</td>
<td>1.24642 1.24642</td>
<td>100.00000 1</td>
<td>1 1</td>
<td>1 1</td>
<td>1 1</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>2</td>
<td>0</td>
<td>1.03043 1.06722</td>
<td>96.55273 0.25003 0.275265 90.83544</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
\begin{table}
\begin{tabular}{|c|c|c|c|c|c|c|}
\hline
n & r1 & r2 & \text{Var}(\mu^*) & \text{Var}(\mu_1^*) & \text{Efficiency} & \text{Var}(\sigma^*) & \text{Var}(\sigma_1^*) & \text{Efficiency} \\
\hline
7 & 2 & 1 & 1.15156 & 1.17961 & 97.62210 & 0.33333 & 0.357092 & 93.34569 \\
7 & 2 & 2 & 1.35547 & 1.37109 & 98.86076 & 0.5 & 0.517885 & 96.54653 \\
7 & 2 & 3 & 1.86149 & 1.86149 & 100.00000 & 0.99992 & 0.99992 & 100.00000 \\
7 & 3 & 0 & 1.32281 & 1.35297 & 97.77083 & 0.33339 & 0.350585 & 95.09534 \\
7 & 3 & 1 & 1.63675 & 1.65116 & 99.12728 & 0.50002 & 0.510907 & 97.86908 \\
7 & 3 & 2 & 2.48982 & 2.48982 & 100.00000 & 1.00008 & 1.00008 & 100.00000 \\
7 & 4 & 0 & 1.90754 & 1.92032 & 99.33449 & 0.5001 & 0.507367 & 98.56770 \\
7 & 4 & 1 & 3.09208 & 3.09208 & 100.00000 & 1 & 1 & 100.00000 \\
8 & 5 & 0 & 3.66228 & 3.66228 & 100.00000 & 1.00040 & 1 & 100.04000 \\
8 & 0 & 1 & 0.73810 & 0.81492 & 90.57331 & 0.16668 & 0.24117 & 69.11307 \\
8 & 0 & 2 & 0.76346 & 0.82553 & 92.48119 & 0.2 & 0.28149 & 71.05048 \\
8 & 0 & 3 & 0.78937 & 0.83284 & 94.78051 & 0.25031 & 0.33916 & 73.80292 \\
8 & 0 & 4 & 0.80972 & 0.83426 & 97.05847 & 0.33332 & 0.42911 & 77.67705 \\
8 & 0 & 5 & 0.82066 & 0.82669 & 99.27059 & 0.5 & 0.59184 & 84.48229 \\
8 & 0 & 6 & 0.82364 & 0.82364 & 100.00000 & 1 & 1 & 100.00000 \\
8 & 1 & 0 & 0.81030 & 0.860119 & 94.20789 & 0.16669 & 0.201503 & 82.72333 \\
8 & 1 & 1 & 0.85503 & 0.898382 & 95.17444 & 0.20002 & 0.236896 & 84.43668 \\
8 & 1 & 2 & 0.90899 & 0.944644 & 96.22567 & 0.24991 & 0.28875 & 86.54892 \\
8 & 1 & 3 & 0.97741 & 1.0037 & 97.38069 & 0.3333 & 0.372682 & 89.44086 \\
8 & 1 & 4 & 1.07267 & 1.08759 & 98.62816 & 0.49998 & 0.534617 & 93.52116 \\
8 & 1 & 5 & 1.24642 & 1.24642 & 100.00000 & 1 & 1 & 100.00000 \\
8 & 2 & 0 & 0.84103 & 0.859098 & 95.61843 & 0.20003 & 0.225188 & 88.82800 \\
8 & 2 & 1 & 1.03043 & 1.06722 & 96.55273 & 0.25003 & 0.275256 & 90.83544 \\
8 & 2 & 2 & 1.15156 & 1.17961 & 97.62210 & 0.33332 & 0.357092 & 93.34569 \\
8 & 2 & 3 & 1.35547 & 1.37109 & 98.86076 & 0.5 & 0.517885 & 96.54653 \\
8 & 2 & 4 & 1.86149 & 1.86149 & 100.00000 & 0.99992 & 0.99992 & 100.00000 \\
8 & 3 & 0 & 1.14898 & 1.18936 & 96.60490 & 0.25006 & 0.269638 & 92.73915 \\
8 & 3 & 1 & 1.32281 & 1.35297 & 97.77083 & 0.33339 & 0.350585 & 95.09534 \\
8 & 3 & 2 & 1.63675 & 1.65116 & 99.12728 & 0.50002 & 0.510907 & 97.86908 \\
8 & 3 & 3 & 2.48982 & 2.48982 & 100.00000 & 1.00008 & 1 & 100.00800 \\
8 & 4 & 0 & 1.48768 & 1.52042 & 97.84665 & 0.33342 & 0.347317 & 95.99876 \\
8 & 4 & 1 & 1.90754 & 1.92032 & 99.33449 & 0.5001 & 0.507367 & 98.56770 \\
8 & 4 & 2 & 3.09208 & 3.09208 & 100.00259 & 1 & 1 & 100.00000 \\
8 & 5 & 0 & 2.16523 & 2.17679 & 99.46894 & 0.50020 & 0.505482 & 98.95506 \\
8 & 5 & 1 & 3.66228 & 3.66228 & 100.00000 & 1.00040 & 1 & 100.04000 \\
8 & 6 & 0 & 4.19737 & 4.19737 & 100.00000 & 1.00040 & 1 & 100.04000 \\
\hline
\end{tabular}
\end{table}

**ACKNOWLEDGEMENT**

I am very grateful to professor Abdul-Samad-Hirai former Director and Professor of Statistics Institute of Statistics, University of Punjab under his kind guidance this work has been completed.
REFERENCES

PREDICTION OF HOUSEHOLD EXPENDITURE ON THE BASIS OF HOUSEHOLD CHARACTERISTICS

Zahoor Ahmad and Abeeda Fatima
Department of Statistics, University of Gujrat, Gujrat, Pakistan
Email: zahoor.ahmed@uog.edu.pk; ezza.fatima@yahoo.com

ABSTRACT

There has been an increasing recognition that some aspects of household (HH) expenditures of potential importance for policymaker. It could potentially affect the effectiveness of policy intervention and may even lead to unintended consequences. The objective of this research is to predict three levels of household expenditures on the basis of household characteristics and to identify the significant household characteristics those can affect the household expenditures. For this purpose, data is obtained from Household Integrated Economic Survey (HIES) 2007-08 conducted by Federal Bureau of Statistics (FBS). In this study, dependent variables is household expenditures (low, medium and high), and independent variables are; household income, assets, number of earners, family size, region, highest education level of household, literacy and sex of household head. Multilayer Perceptron Neural Network has been used for data analysis. Results showed that all independent variables are playing a significant role to predict three levels of household expenditures. The order of importance of independent variables is HH income, HH assets, number of earners per HH, HH size, sex of HH head, region, literacy of HH head, and highest education level in household.

KEY WORDS

Household Expenditures, Multilayer Perceptron Neural Network.

1. INTRODUCTION

Income is the consumption and savings opportunity acquired by an entity within a specified time frame, which is generally expressed in monetary terms. However, for households and individuals, “income is the sum of all the wages, salaries, profits, rents, interests’ payments and other forms of earnings received in a given period of time”.

Household income is habitually a main determinant of household expenditure patterns. The differences between expenditure patterns are largely an indication of differences in income between household groups or individual households. There are many factors which are influencing the preference on one hand and the income levels on the other hand, and they both are affecting the expenditure patterns. Expenditure patterns of household do not always differ spectacularly, particularly when controlling for income [Punt, et al. (2003)].

While analyzing the household income and consumption expenditures of household, we have to reflect on the sources of income. In Pakistan, there are many sources of income (such as wage and salaries, property owner occupied dwelling, livestock, crop, non-agricultural activities, social insurance benefits which are including gift and assistance, pension, domestic remittances, foreign remittances and other income).
There is a highly positive relationship between income share from rural non-agricultural activities and rising per capita income and a negative relationship between share of agricultural production and per capita income. As household income increases, the propensity to move away from agricultural production to the non-agricultural activities is a pattern similar to Engel’s Law, which assumed that poorer households spend more share of their income on food than wealthier households [Akita (1999)].

Family size and dependency are highly correlated with expenditure and considered as important poverty predictors. Proportion of members greater than 65 and children are representing dependency. In rural areas, land, poultry, ownership of livestock, non-residential and residential property are high positively related to household expenditure. Further, owner cultivator, medium and large farmers (land ownership of 12.5 acres and greater) play a vital role in distinguishing between non-poor from poor.

There are some factors which are creating disparity in the earnings including regional location factors, sex, sector of employment, marriage, and some other characteristics. Mostly youngest and oldest age groups contribute more to overall inequality in the earnings. [Nasir and Mahmood (1998)]. The nature of inequality has varied between income and consumption expenditure. The income inequalities is in the middle class households while consumption inequality is low within middle and lower income group as compared to upper expenditure class [Blacklow and ray (1999)].

In Pakistan, there is a substantial gender differences exist between intra household allocations of educational expenditures such that priority has given to males over females [Aslam (2008)]. In particular, the consumption on some investment categories (like education, health, durable goods, patrimony and savings) showed to be significantly larger for those households who are receiving remittances (external or internal) than for non remittance households. On the average, household receiving external remittance have more income than those who are receiving internal remittances or non remittance households [Rivera and González (2009)].

Receiving remittances and the sex of the household head together affect expenditure allocations of household. Female-headed households spend most of their expenditures on food and education, and low expenditures on durable and consumer goods, housing and other items. Female headed households are not homogeneous in the sense that those households who are receiving remittances from abroad consume considerably more of their expenditures on durable and consumer goods and housing than those who are receiving remittances from within Ghana or not receiving any kind of remittances.

Remittances have insignificant effect on households’ education expenditures. Internal remittances tend to decrease the proportion of expenditures devoted to investments in agriculture and housing. Furthermore, they do not have significant impact on education and other expenditures. This could be construed as evidence of a non-productive use of internal remittances, in the same way as external remittances. However, a strong positive effect on health expenditures is observed. [Clement (2011)].

The gender, location and region of household head effects expenditure patterns. Female headed households consume more on food, fuel and clothing while male headed households consume more on income tax, housing, transport, savings, services and other items including beverages. There are differences in expenditure patterns between households from different income groups and racial groups. [Punt, et al. (2003)].
2. OBJECTIVE OF STUDY

- To predict three level of household expenditures on the basis of HH characteristics.
- To identify the significant household characteristics those can affect the HH expenditures

3. LITERATURE REVIEW

Akita (1993) conducted a research to explore the factors underlying income inequality in Indonesia by using household data from 1987, 1990 and 1993 national socio economic survey. This was done by using Theil inequality decomposition technique. Results showed that was that inter provincial inequality has not been played an important role in inequality while education and gender inequality were a significant determinant of expenditure inequality. Household expenditure increases as the age of household head increase, but after age of 45 it again decreases. Household size also tended to increases with the age of household head but after the children become independent, again it became small. Income inequality within age group also had a tendency to increases with the age of household head. As household size increases, household expenditure increases while per capita expenditures tended to decrease.

Agnes and John (2000) was conducted a research to judge against household’s unitary models and collective models by using household data from Indonesia, Bangladesh, Ethiopia, and South Africa. They were presented measures of individual characteristics that were greatly correlated to the bargaining power, assessed at the time of marriage. In addition, the collective model predicted that intra household allocations reflecting the differences between "bargaining power" and preferences of household members within the household. In this study unitary model was rejected. Results suggested that assets which were controlled by women positively and significantly affected household expenditure allocations toward the subsequent generation, such as clothing and education of children. They also examined that parents had distinguishable preferences to their sons and daughters within and across all countries.

Kirkpatrick and Tarasuk (2003) conducted a research to compare food expenditure patterns of households on the basis of their income level and to observe the relationship between food expenditure patterns and the absence or presence of housing payments amongst low income households by using chi-square and multiple regression analysis. Data is obtained from Family Food Expenditure Survey-1996 conducted by Statistics Canada. Their findings indicated that access to fruits, vegetables and milk products among households may possibly be constrained in the situation of low incomes. This study suggested that there is a greater need to pay attention to the affordability of nutritious foods for low income groups.

Punt, et al. (2003) conducted a study to analyze the impact of different levels of household income on household expenditure patterns and to know how the expenditure patterns differ among households of different income level from different region and ethnic groups. In his study latest household income and expenditure survey (IES) data 1995 was used. The descriptive analysis was performed. It was found that the levels of household income are most important factor to determine household expenditure patterns as well as various other factors like race and location of households also affect expenditure patterns.
Guzman (2006) conducted a research to investigate how household budget allocation affected by remittances, sex of remitter and sex of household head by using data from Ghana Living Standard Survey 1998-99. Descriptive analysis and fractional logit regression analysis was used. The results indicated that all factors separately and jointly affected household expenditure allocations.

Lemay (2006) conducted a research to analysis of the relationship between household expenditures and the allocation of income within the household and to test which model (cooperative model or non-cooperative model) best suited household’s consumption decision process by using data from Benin. Two stage OLS regression was applied. Results were rejected the cooperative model and suggested that a non-cooperative model was best matched for describing interaction between members of household. It was also observed that African women significantly expend more share of their income in socially desirable manners while men spend considerably more on personal expenditure.

Aslam (2008) conducted a research to investigate the gender discrimination in the intra-household allocation of educational expenditure by using data from Pakistan Integrated Household Survey (PIHS) 2001-02. Engel curves and Hurdle models were estimated to know about the gender bias in the decision to enroll in school or in spending educational expenditures. It was observed that a pro-male bias plays an important role in both decisions at junior and secondary school ages. However at primary school age, both face discrimination in decision about enrolment in school only. It is also suggested that gender bias in educational expenditures is within household but not across household phenomenon.

Morrison, et al. (2008) conducted a research to know about the relationship between gender and remittances on expenditure patterns of household in Ghana by using data from Ghana Living Standard Survey round four (GLSS 4) 1998/99. Fractional logit regression was used. Results showed that female-headed households spend more percentages on food and education while smaller percentage of their expenditure on durable goods, consumer goods, housing and other goods. But on the other hand those female-headed households which were receiving international remittances consume a smaller share on food than do female-headed households that not receive these remittances. Households receiving remittances from female remitters were allocated a greater share on their health and other goods while a smaller share on food.

4. DATA SOURCE AND ANALYSIS

The data for this study is taken from the Household Integrated Economics Survey (HIES) 2007-08 which conducted by Federal Bureau of Statistics (FBS), Pakistan. There are some missing cases related to variables: expenditure (low, medium and high), income, assets, number of rooms, number of earners, age of head, family size, proportion of members less than 5, proportion of out of school children, region of head, proportion of members greater than 65 of households. So we have used only eligible cases. Ordinal Logistic Regression and Multilayer Perceptron Neural Network have been used for data analysis.

5. RESULTS AND DISCUSSION

Multilayer Perceptron Neural Network has been used for data analysis. Table A-1 represents the network summary analysis 1043(68.3%) cases assigned to the analysis
sample and 485(31.7%) cases to the holdout sample. Total no. of valid cases in this analysis are 1528(100%). Here model excluded 14104 cases from the analysis due to missing which are not match with the independent variables while total no of cases are 15632.

Table A-2 shows the network information, total no of units used by the model in input layer are13 and no of units in the hidden layers is 5 and units in output layer are 3. In this network model include total 8 independent variables in which 4 represent as factors and remaining 4 are covariates. Standardized rescaling method is used for input layers and softmax activation function used in output layers and the cross entropy error function used.

Figure A-1 showed that, the category one (low expenditures) has high positive relationship with 1, 4 and 5 units of first hidden layer in this model and negative relation exists with 2 and 3 units of hidden layer. The second category (medium expenditures) show strong positive relationship with 2 and 5 units in the hidden layers and vary low positive relationship with 1 and 3 units of hidden layer and strong negative relationship with 4 unit of hidden layer. 3rd category (high expenditures) show positive relationship with 2 and strong negative relationship with 1, 4 and 5 hidden layers units and having very low negative relationship with 3 unit of hidden layer this is same situation shown in table of parameter estimation of network model (table A-4).

Table A-3 represents the model summary of training sample and holdout sample which connected to error and incorrect prediction of the response variable household expenditures. Value of cross entropy error functions in this model is 348.09 which are very low corresponds to all sample data and this indicated that our estimated model is better. Training sample contains 14.1% incorrect prediction and in holdout incorrect prediction is 16.3%.

Table A-5 represents the summary of classification results of network model. Cells on the main diagonal showed the correct classifications and on off diagonal of the cross-classification are showed incorrect classifications of the response category. In training sample 189 cases are correctly classify in low category of expenditure who are previously fall in this category and 291 cases are correctly classify in medium expenditure category as they previously fall in this category and 71 cases incorrectly specify in other two category. 416 cases are correctly classified in high expenditure category and 40 cases incorrectly classify in other two categories. So overall correctly classification in training sample for all three categories is 85.9%. In holdout sample 79 cases correctly classify in low expenditure category and 20 cases are incorrectly classified in other two categories. 146 cases are correctly classified in medium category and 40 cases incorrectly specify in other two categories. 181 cases correctly classify in high category and 19 cases incorrectly classify in other two category. Overall correctly classification in holdout sample for all three categories is 83.7%.

For categorical response variable, the predicted-by-observed chart demonstrate clustered of box plots with their predicted pseudo-probabilities for the both the training and testing samples. It is an alternative way to understand the correct classification for the purpose to check realistic importance of Neural Network. On the x-axis this showed the observed response categories, and the legend corresponding to the predicted categories. The part of box plot above the 0.5 on y-axis identified the correct predictions and the part below 0.5 represents incorrect predictions of response categories.
Figure A-2 show the predicted -by-observed chart most left box plot shows, for cases that have observed category for low expenditure but shows the predicted pseudo-probability of category for low in this category. By looking this chart we can say that all portion of box plot is above 0.5 this indicates that predicted pseudo probabilities are very high for low category and box plot for all remaining two categories are below 0.5 and incorrectly predicted in low category. In medium expenditure category the middle plot show all the portion of this box plot above the 0.5 so its predicted pseudo probability are high in medium category and all other categories of low and high expenditure are incorrectly classified in medium category. The next category of high expenditures the portion of last box plot is above 0.5 so it’s predicted probability high and correctly classify in high expenditure category and remaining two categories are incorrectly classified in high category of expenditure.

Figure A-3 shows the receiver OC curve (ROC) and three curves; 1st for the low expenditures category, 2nd for the category of medium expenditures and 3rd category for high expenditures. The curves of these three categories are very close to upper left corner of the plot; which indicating that our model is better fit for given data and area covers for the category low expenditures is 0.976, area under the curve covers for the medium expenditures is .937and area covers for the category for high expenditures is 0.979 see more in table A-6. By looking this table we can say that mostly area covered closed to 1 which indicates perfect fit on 1.

In Figure A-4, the cumulative gains chart that is displayed the percentage of the overall numbers of cases in a given category is “gained” by targeting a percentage of the total number of cases. In this case, very first point on the curve against the low expenditures category is at (10%, 48%), which tell us that if you score a data set with the network and sort out all of the cases by predicted pseudo-probability of low expenditures category. This indicates that, the top 10% to contain approximately 48% of all of the cases that actually gained in category low expenditures.

Figure A-5 shows the lift chart that is derived from the cumulative gain chart; the values on the y axis correspond to the ratio of the cumulative gain for each curve to the baseline. Thus, the lift at 10% for the category 48% / 10% = 4.8 it means approximately 4 cases correctly specify at top 10% of low expenditures.

Table A-7 and also the Figure A-6 shows the importance of independent variables In this network model, 4 variables are playing very important role in predicting the expenditure categories of low, medium and high. This network model gives highest importance to Household (HH) income (100) than all others and second importance gives to household (HH) assets (72.3%) third important variable is no of earners per household (28.8%), forth importance give to household size (25.4%). In the light of all above results related to importance of predictor we conclude that all that play a significant in predicting the expenditures level.

6. CONCLUSION

Results shows that household income, assets, number of earners, family size, region and sex of household head are significant variables those can best discriminate the respondents among levels (low, medium and high) of household expenditure while
highest education level and literacy are insignificant variables. These significant variables
can help to predict the respondents among the levels of household expenditures.

Results from ordinal logistic regression shows that in first category of expenditures
75.3% cases, in second category, 78.8% cases while in last category 86.7% cases are
correctly classified and overall correct classification for all three categories is 80.26%.

Neural network shows that four variables are playing most important role in
predicting the expenditure categories of low, medium and high. The network has given
highest importance to HH income rather than all others, second importance is given to
HH assets, third important variable is number of earners per household, and forth
importance is given to household size. Overall correct classification by the network for
all three categories is 83.7%.

REFERENCES

   Relations: New Empirical Evidence from Four Developing Countries. International
   Food Policy Research Institute, FCND discussion paper 84, Washington DC.
   of Household Expenditures in Indonesia: A Theil Decomposition Analysis. The
   Developing Economies, 37(2), 197-221.
   Applied Economics, 40(20), 2573-2592.
   Inequality Estimates: The Australian Evidence, 1975/76 to 1993/94. The Australian
   and Gender on Household Expenditure Patterns: Evidence from Ghana. Gender and
   Development (PRMGE), World Bank.
   Household Food Expenditure Patterns in Canada. Public Health Nutrition, 6(6), 589-597.
    Women. Co-Publication of World Bank and Palgrave Macmillan, 125-152.
12. Punt, C., Pauw, K., Mohube, E., Gilimani, B., Rantho, L., Leaver, R., McDonald, S.,
    1995. PROVIDE Project Background Paper, 2, 1-16.
    Expenditure Patterns of Rural Mexico. Unpublished paper.
APPENDIX-A

### Table A-1: Case Processing Summary

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sample Training Holdout</td>
<td>1043</td>
<td>68.3%</td>
</tr>
<tr>
<td>Valid</td>
<td>1528</td>
<td>100.0%</td>
</tr>
<tr>
<td>Excluded</td>
<td>14104</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>15632</td>
<td></td>
</tr>
</tbody>
</table>

### Table A-2: Network Information

<table>
<thead>
<tr>
<th>Input Layer</th>
<th>Factors</th>
<th>1</th>
<th>Sex of head</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>2</td>
<td>Region</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3</td>
<td>Literacy</td>
</tr>
<tr>
<td></td>
<td></td>
<td>4</td>
<td>Highest education in HH</td>
</tr>
<tr>
<td>Covariates</td>
<td></td>
<td>1</td>
<td>Household size</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2</td>
<td>No of earners per HH</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3</td>
<td>HH assets</td>
</tr>
<tr>
<td></td>
<td></td>
<td>4</td>
<td>HH income</td>
</tr>
<tr>
<td>Number of Units</td>
<td>13</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rescaling Method for Covariates</td>
<td>Standardized</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hidden Layer(s)</td>
<td>Number of Hidden Layers</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Number of Units in Hidden Layer 1</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Activation Function</td>
<td>Hyperbolic tangent</td>
<td></td>
</tr>
<tr>
<td>Output Layer</td>
<td>Dependent Variables</td>
<td>1</td>
<td>Expenditures</td>
</tr>
<tr>
<td></td>
<td>Number of Units</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Activation Function</td>
<td>Softmax</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Error Function</td>
<td>Cross-entropy</td>
<td></td>
</tr>
</tbody>
</table>

a. Excluding the bias unit

### Table A-3: Model Summary

<table>
<thead>
<tr>
<th></th>
<th>Cross Entropy Error</th>
<th>Percent Incorrect Predictions</th>
<th>Stopping Rule Used</th>
</tr>
</thead>
<tbody>
<tr>
<td>Training</td>
<td>348.097</td>
<td>14.1%</td>
<td>Maximum number of epochs (100) exceeded 00:00:01.716</td>
</tr>
<tr>
<td>Holdout</td>
<td>Percent Incorrect Predictions</td>
<td>16.3%</td>
<td></td>
</tr>
</tbody>
</table>

Dependent Variable: Expenditures
### Table A-4: Parameter Estimates

<table>
<thead>
<tr>
<th>Predictor</th>
<th>H(1:1)</th>
<th>H(1:2)</th>
<th>H(1:3)</th>
<th>H(1:4)</th>
<th>H(1:5)</th>
<th>Output Layer</th>
</tr>
</thead>
<tbody>
<tr>
<td>Input Layer (Bias)</td>
<td>-0.041</td>
<td>1.870</td>
<td>0.189</td>
<td>0.174</td>
<td>-1.176</td>
<td>Low</td>
</tr>
<tr>
<td>[sex=1]</td>
<td>0.696</td>
<td>1.117</td>
<td>0.350</td>
<td>0.498</td>
<td>-0.957</td>
<td></td>
</tr>
<tr>
<td>[sex=2]</td>
<td>-1.019</td>
<td>1.101</td>
<td>0.306</td>
<td>0.021</td>
<td>-2.288</td>
<td></td>
</tr>
<tr>
<td>[region=1]</td>
<td>-0.315</td>
<td>0.451</td>
<td>0.574</td>
<td>0.222</td>
<td>-0.031</td>
<td></td>
</tr>
<tr>
<td>[region=2]</td>
<td>0.318</td>
<td>0.421</td>
<td>0.325</td>
<td>0.409</td>
<td>-0.167</td>
<td></td>
</tr>
<tr>
<td>[literacy=0]</td>
<td>0.897</td>
<td>0.307</td>
<td>0.618</td>
<td>-0.192</td>
<td>-0.502</td>
<td></td>
</tr>
<tr>
<td>[literacy=1]</td>
<td>-0.019</td>
<td>0.362</td>
<td>0.601</td>
<td>-0.096</td>
<td>-0.274</td>
<td></td>
</tr>
<tr>
<td>[edu_categories=1]</td>
<td>-0.126</td>
<td>0.476</td>
<td>0.394</td>
<td>-0.281</td>
<td>-0.187</td>
<td></td>
</tr>
<tr>
<td>[edu_categories=2]</td>
<td>0.032</td>
<td>0.631</td>
<td>0.081</td>
<td>-0.685</td>
<td>-0.192</td>
<td></td>
</tr>
<tr>
<td>[edu_categories=3]</td>
<td>-0.164</td>
<td>0.808</td>
<td>-0.495</td>
<td>1.019</td>
<td>-0.633</td>
<td></td>
</tr>
<tr>
<td>HH size</td>
<td>-0.606</td>
<td>0.211</td>
<td>0.180</td>
<td>1.165</td>
<td>0.008</td>
<td></td>
</tr>
<tr>
<td>No of earners</td>
<td>0.476</td>
<td>-0.163</td>
<td>1.587</td>
<td>-0.094</td>
<td>0.000</td>
<td></td>
</tr>
<tr>
<td>HH assets</td>
<td>-2.822</td>
<td>7.911</td>
<td>0.000</td>
<td>1.617</td>
<td>-6.372</td>
<td></td>
</tr>
<tr>
<td>HH income</td>
<td>-2.278</td>
<td>2.189</td>
<td>2.266</td>
<td>-5.763</td>
<td>-2.154</td>
<td></td>
</tr>
</tbody>
</table>

| Hidden Layer 1 (Bias) | .615 | 2.218 | -2.571 | H(1:1) | 2.056 | 0.556 | -2.982 |
| H(1:2) | -5.431 | 2.033 | 3.143 | H(1:3) | -0.488 | 0.480 | -0.190 |
| H(1:4) | 3.239 | -1.487 | -2.221 |
| H(1:5) | 1.697 | 2.721 | -5.499 |

### Table A-5: Summary Classification

<table>
<thead>
<tr>
<th>Sample</th>
<th>Observed</th>
<th>Predicted low</th>
<th>Predicted Medium</th>
<th>Predicted high</th>
<th>Percent Correct</th>
</tr>
</thead>
<tbody>
<tr>
<td>Training</td>
<td>Low</td>
<td>189</td>
<td>33</td>
<td>3</td>
<td>84.0%</td>
</tr>
<tr>
<td></td>
<td>medium</td>
<td>30</td>
<td>291</td>
<td>41</td>
<td>80.4%</td>
</tr>
<tr>
<td></td>
<td>High</td>
<td>0</td>
<td>40</td>
<td>416</td>
<td>91.2%</td>
</tr>
<tr>
<td></td>
<td>Overall</td>
<td>21.0%</td>
<td>34.9%</td>
<td>44.1%</td>
<td>85.9%</td>
</tr>
<tr>
<td>Holdout</td>
<td>Low</td>
<td>79</td>
<td>18</td>
<td>2</td>
<td>79.8%</td>
</tr>
<tr>
<td></td>
<td>medium</td>
<td>12</td>
<td>146</td>
<td>28</td>
<td>78.5%</td>
</tr>
<tr>
<td></td>
<td>High</td>
<td>0</td>
<td>19</td>
<td>181</td>
<td>90.5%</td>
</tr>
<tr>
<td></td>
<td>Overall</td>
<td>18.8%</td>
<td>37.7%</td>
<td>43.5%</td>
<td>83.7%</td>
</tr>
</tbody>
</table>

Dependent Variable: Expenditures

### Table A-6: Area under the Curve

<table>
<thead>
<tr>
<th></th>
<th>Area</th>
</tr>
</thead>
<tbody>
<tr>
<td>Expenditures</td>
<td>.976</td>
</tr>
<tr>
<td>Low medium high</td>
<td>.937</td>
</tr>
<tr>
<td></td>
<td>.979</td>
</tr>
<tr>
<td></td>
<td>Importance</td>
</tr>
<tr>
<td>--------------------------</td>
<td>------------</td>
</tr>
<tr>
<td>Sex of head</td>
<td>.022</td>
</tr>
<tr>
<td>Region</td>
<td>.016</td>
</tr>
<tr>
<td>Literacy</td>
<td>.014</td>
</tr>
<tr>
<td>Highest education in HH</td>
<td>.031</td>
</tr>
<tr>
<td>household size</td>
<td>.103</td>
</tr>
<tr>
<td>no of earners per HH</td>
<td>.117</td>
</tr>
<tr>
<td>HH assets</td>
<td>.293</td>
</tr>
<tr>
<td>HH income</td>
<td>.405</td>
</tr>
</tbody>
</table>
Fig. A-1: Network Diagram
Prediction of household expenditure on the basis of…

Fig. A-2: Predicted by Observed Chart

Fig. A-3: ROC Curve
Fig. A-4: Cumulative Gains Chart

Fig. A-5: Lift Chart
Fig. A-6: Independent Variable Importance Chart
**APPENDIX-B**

**Prediction of HH expenditures by Using MLP Model Manually**

To estimate or predict the value form the MLP model, original values of the independent variables and weights estimated by the model is required. As the current model have three layers; input layer, hidden layer and output layer. So in first step hidden layer receives the weighted sum of incoming signals (information from independent variables) sent by the input layer and processes on it in the hidden layer by means of an activation function. In the existing model hyperbolic tangent activation function is used for hidden layer. The form of hyperbolic tangent activation is \( f(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} \). The hidden layer in turn sends an output signal towards the neurons in the next layer that is an output layer. So processing on the weighted information received by the output layer is done on the same way as in hidden layer that is by applying an activation function on it. Here softmax activation function has been used by the model which has the form:

\[
\gamma(c_k) = \frac{e^{c_k}}{\sum e^{c_j}}.
\]

Detailed manual calculation for each layer is given below.

**Estimation of HH expenditures by Using MLP Model:**

As to obtain the value from the model, information on independent variables and weights are required. The table given below is providing the information on each indicator and weights are given in table A-4.

<table>
<thead>
<tr>
<th>Indicators</th>
<th>Respondent data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sex of HH head</td>
<td>1</td>
</tr>
<tr>
<td>Region of HH</td>
<td>1</td>
</tr>
<tr>
<td>Literacy of HH head</td>
<td>1</td>
</tr>
<tr>
<td>Highest education level in HH</td>
<td>3</td>
</tr>
<tr>
<td>HH Size</td>
<td>8</td>
</tr>
<tr>
<td>No. of earners per HH</td>
<td>2</td>
</tr>
<tr>
<td>HH Assets</td>
<td>100000</td>
</tr>
<tr>
<td>HH Income</td>
<td>180000</td>
</tr>
</tbody>
</table>

**Calculations for input to hidden layer**

\[
y_j = \sum_{i=1}^{n} w_{ij}x_i \quad \quad i = 1, 2, ..., n
\]

where \( n \) is the number of input variables \( w_{ij} \) denotes the weights for \( ith \) variable and \( x_i \) is the \( ith \) variable.
Prediction of household expenditure on the basis of...

\[ y_1 = bias + w_{11}x_1 + w_{21}x_2 + w_{31}x_3 + w_{41}x_4 + w_{51}x_5 + w_{61}x_6 + w_{71}x_7 + w_{81}x_8 \]
\[ = -0.040641028 + (0.696263)(1) + (-0.315)(1) + (-0.01873)(1) + (-0.16357)(3) + (-0.60623)(8) + (0.475649)(2) \]
\[ + (-2.82156)(100000) + (-2.27777)(180000) \]
\[ = -692157.87 \]

\[ y_2 = bias + w_{12}x_1 + w_{22}x_2 + w_{32}x_3 + w_{42}x_4 + w_{52}x_5 + w_{62}x_6 + w_{72}x_7 + w_{82}x_8 \]
\[ = 1.869862 + (1.869862)(1) + (0.451)(1) + (0.361979)(1) + (0.808095)(3) + (0.211302)(8) + (-0.16294)(2) \]
\[ + (7.910957)(100000) + (2.188577)(180000) \]
\[ = 1185048 \]

\[ y_3 = bias + w_{13}x_1 + w_{23}x_2 + w_{33}x_3 + w_{43}x_4 + w_{53}x_5 + w_{63}x_6 + w_{73}x_7 + w_{83}x_8 \]
\[ = 0.188663 + (0.188663)(1) + (0.574)(1) + (0.60058)(1) + (-0.49491)(3) + (0.180176)(8) + (1.586747)(2) \]
\[ + (-0.00044)(100000) + (2.265691)(180000) \]
\[ = 407784.7 \]

\[ y_4 = bias + w_{14}x_1 + w_{24}x_2 + w_{34}x_3 + w_{44}x_4 + w_{54}x_5 + w_{64}x_6 + w_{74}x_7 + w_{84}x_8 \]
\[ = 0.17432 + (0.17432)(1) + (0.222)(1) + (-0.09552)(1) + (1.018984)(3) + (1.164808)(8) + (-0.09366)(2) \]
\[ + (1.617084)(100000) + (-5.7626)(180000) \]
\[ = -875547 \]

\[ y_5 = bias + w_{15}x_1 + w_{25}x_2 + w_{35}x_3 + w_{45}x_4 + w_{55}x_5 + w_{65}x_6 + w_{75}x_7 + w_{85}x_8 \]
\[ = -1.17607 + (-1.17607)(1) + (-0.031)(1) + (-0.27371)(1) + (-0.63291)(3) + (0.008353)(8) + (0.000323)(2) \]
\[ + (-6.3724)(100000) + (-2.15409)(180000) \]
\[ = -1024981 \]

**Calculation for hidden to output layer**

As the activation function for hidden layer is hyperbolic tangent, so

\[ \tanh(y_1) = \tanh(-692157.8757) = -1 \]
\[ \tanh(y_2) = \tanh(1185048) = 1 \]
\[ \tanh(y_3) = \tanh(407784.7) = 1 \]
\[ \tanh(y_4) = \tanh(-875547) = -1 \]
\[ \tanh(y_5) = \tanh(-1024981) = -1 \]
Calculation for Output layer

\[ z = \text{bias} + w_{16} \tanh(y_1) + w_{26} \tanh(y_2) + w_{36} \tanh(y_3) + w_{46} \tanh(y_4) + w_{56} \tanh(y_5) \]

\[ = 0.615329 + (2.055585)(1) + (-0.031)(1) + (-0.27371)(1) \]

\[ + (-0.63291)(-1) + (0.008353)(-1) \]

\[ = 0.99 \]

Calculation for Activation function for output layer

\[ \gamma(c_k) = \frac{e^{c_k}}{\sum e^{c_j}} \]

\[ \gamma(-12.29417056) = \frac{0.00000046}{65139.7800044} = 0.00000001 \]

\[ \gamma(2.940932) = \frac{18.93348}{65139.7800044} = 0.000291 \]

\[ \gamma(11.084) = \frac{65120.85}{65139.7800044} = 0.999709 \]

As the probability is high in the third category, so we can say that this respondent has high expenditures.
CLASSIFICATION OF HOUSEHOLDS WITH RESPECT TO POVERTY
BY USING CLUSTER ANALYSIS

Zahoor Ahmad and Zainab Ejaz
Department of Statistics, University of Gujrat, Gujrat, Pakistan
Email: zahoor.ahmed@uog.edu.pk; zainab.ejaz40@gmail.com

ABSTRACT

Poverty has many dimensions in Pakistan. The major challenge of today is poverty reduction. Poverty reduction has been a major objective of many developing countries. The main objective of this paper is to explore the factors which are playing a significant role for the cluster of non poor household and poor household and also investigate the direction and significance of importance variables. The independent variables are household income, family size, sex ratio, household education, dependency ratio. Data was obtained from Pakistan Social and Living Standard Measurement Survey (PSLM) 2008-2009. Two steps cluster analysis are used for data analysis purpose. Results of this paper demonstrates that household education and household income play significant role in reducing poverty for non-poor households, while dependency ratio and family size play negative role in non-poor household. But in poor household dependency ratio and family size play positive role. It means that poverty increases due to increasing the number of dependents and family size. Poverty in household can be reduced by lowering the household size and decreasing number of dependants.

KEY WORDS

Poverty; Two step cluster analysis.

1. INTRODUCTION

Poverty is a widespread problem that particularly afflicts developing countries. According to the World Bank (2009) report, people living below the poverty line (based on $2-a-day criterion) account for more than 80 percent of the population in India, Bangladesh and Nepal, 73.6 percent in Pakistan, and 41.6 percent in Sri Lanka. In a country like Pakistan where nearly 40% people (based on one $ per day income) live below poverty line and per capita income is less then 1000$. Poverty has many dimensions in Pakistan. People have not only low incomes but they also are suffering from lack of access over basic needs. The major challenge of today is poverty reduction.

Poverty cannot be described it can only be felt. One knows more about poverty when he is hungry and cannot purchase food, he and his children want new clothes but they can’t purchase it because of low income, he’s sick and doesn’t have money to have medicine, he wants to send his children to school but can’t bear educational expenditures. The world Development Reports define poverty as “pronounced deprivation in well being”.

369
Education is considered as an essential determinant of household poverty. The probability of being a poor household decreases with higher the level of education in a household [Chudhary (2009)]. Janjua and Kemal (2011) find out that income growth play an important role in poverty reduction, while income distribution does not play important role in poverty reduction. Large family size is considered an important contributor to household poverty and Vulnerability to Poverty, as increase in family size there are more chances household being a poor [Orbeta (2005)].

Jamal (2005) conducted a study to find out the predictor of poverty and their estimates in Pakistan. The log of total household expenditures was used as dependent variable. The data was analyzed using multiple regression analysis. The finding showed that household education level, number of dependents, household head, domestic and overseas transfer were major determinant of poverty which classify the individual into poor and non poor category. But in rural areas population of livestock, ownership of land, played an important role in categorize the individual into poor and non poor category.

Bogale, at al. (2005) identified the determinant of rural poverty in Ethiopia. The data was analyzed by Logit model with binary dependent variable (probability of being poor household). Independent variables were age, education level, dependency ratio, per capita expenditure, per capita income, gender of household head, landholding, population of livestock and number of oxen owned. The result revealed that rural poverty was strongly associated with lack of human capital and household assets like ownership of land.

Orbeta (2005) conducted a study to determine relationship between the family size and household poverty and Vulnerability in Philipines. Data was analyzed by using Multivariate and cross tabulation analysis. The finding of this study demonstrated that family size was important determinant of household poverty. Per capita income, expenditures savings adversely affected as family size increase. Bigger family size was associated with higher vulnerability to poverty.

Rehman (2006) conducted a study to determine impact of education level and employment on poverty reduction among the rural household in Bangladesh. The objectives of this study were, to access the poor and the non poor to primary and secondary level of education and their impact on the poverty reduction, and also the link between the poverty and rate of secondary school completion and success rates secondary school. The data of this study was based on school level and household survey, which was conducted under the programme of research on chronic poverty. Data was analyzed using Multiple regression analysis .The result of this study found that there was huge disparities between the poor and the non poor groups with respect to education level from primary to secondary to SSC completion and level of education was positively associated with salaried employed.

Chaudhry, et al. (2009) argues that socioeconomic and demographic characteristics have an impact on household poverty. He concluded that Dependency ratio, female head household and residence in katcha house play significant role with the probability of being poor household. while Household size, education levels, participation rate, female and male ratio of the workers, the dependency ratio, persons per room, population livestock per household, landholding per household, age of the household head were significantly impact on household per capita income.
Ahmad and Zainab (2009) find out the factors which affect the rural poverty in southern Punjab (Pakistan). Data was analyzed by employing the logit regression model, with dummy or categorical variable as dependent variable, which was defined by the adjusted official poverty line (the rural poverty line in Rs. 865.70 per adult per month). The result showed that household size, education levels, participation rate, female and male ratio of the workers, the dependency ratio, persons per room, population livestock, landholding, age of the household head, physicals assets were significant associated with poverty.

Akerele and Adewuyi (2011) conducted a study to determine household poverty and socio economic determinant of welfare among household in Ekiti State, Nigeria. The objective of the study was to examine the poverty status of household by using the demographic characteristics. Data was analyzed by using descriptive statistics, poverty index, and multiple regression analysis. Household per capita expenditure used as the dependent variable. The result concluded that female headed households were more susceptible to income poverty as compare to male headed household. Highest levels of poverty were found among those household where 7-9 dependents. Educational levels, gender of household head and dependency ratio were major determinant that exact significance influence on household welfare.

Awan, et al. (2011) conducted a study to estimate the impact of human capital on poverty reduction. Education was very important component of human capital. Here the dependent variable was poverty status which was defined by using the adjusting official poverty line. The binomial logistic regression model was used for estimating the impact of human capital on poverty. The result showed that education and experience was negatively related to poverty status of individual, especially primary and middle education help to reduce the poverty, and suggested that with increase the education level there were greater chances to go out of poverty.

Keeping in view some of the most significant literature review, we come to the factors which are used in this study. We have used different variables that effects poverty using micro data from PSLM 2008-2009. These factors include:

**Household Education Level:** This variable is constructed based on sum of the points (0 to household member having no education, 5 points up to secondary level and 10 points up to college university) divided by household size of concerned household.

**Household Income:** Income represents a very important factor when characterizing the poor. The level of income is important not only for the households, but its distribution among household member and various socioeconomic group.

**Family Size:** Family size is also important factor which effect poverty. Large family size is considered an important contributor to household poverty and Vulnerability to Poverty, as increase in family size there are more chances household being a poor [Orbeta (2005)].

**Dependency Ratio:** For a given household size, a larger number of children and elderly members would imply a smaller number of earners in the household. In the present analysis, the dependency ratio is calculated as the ratio of the number of members below 15 and over 64 to other household members. This ratio allows us to measure the burden on members of the labor force within the household.
Sex Ratio: It is called males-females ratio. This variable is constructed as numbers of males divided by number of females of concerned household.

The dependent variable which is dichotomous has been derived from official poverty line of Government of Pakistan Per Capita per Month Rs.848.79.

2. OBJECTIVE OF THE STUDY

1. To explore the factors which are playing a significant role for the cluster of non poor household and poor household
2. To investigate the direction and significance of important variables.

3. DATA SOURCE AND ANALYSIS

In this study data is obtained from Pakistan Social and Living Standard Measurement Survey 2008-2009 of 75192 cases. We have used five variables that are household education, household income, family size, dependency ratio, sex ratio. Two Step Cluster Analysis was used for data analysis purpose.

4. CLUSTER ANALYSIS

Cluster analysis is a group of multivariate techniques whose primary purpose is to group objects based on the characteristics they possess. It has been referred to as Q analysis typology construction, classification analysis and numerical taxonomy. Cluster analysis classifies objects (e.g., respondents, products, or other entities) so that each object is similar to others in the cluster based on a set of selected characteristics. The resulting clusters of objects should exhibit high internal (within-cluster) homogeneity and high external (between-cluster) heterogeneity. Thus, if the classification is successful, the objects within clusters will be close together when plotted geometrically, and different clusters will be far apart.

In cluster analysis, the concept of variate is again a central issue. The cluster variate is a set of variables representing the characteristics used to compare objects in the cluster analysis. Because the cluster variate includes only the variables used to compare objects, it determines the character of the objects. The focus of the cluster analysis is on the comparison of objects based on the variate, not on the estimation of the variate itself.

In many instances, however, the grouping of objects is actually a means to an end in terms of a conceptually defined goal. The more common roles cluster analysis can play in conceptual development including the following:

1. Data Reduction: A researcher who has collected data by means of a questionnaire may be faced with a large number of observations that are meaningless unless classified into manageable groups. Cluster analysis can perform this data reduction procedure objectively by reducing the information from an entire population or sample to information about specific, smaller subgroups.

2. Hypothesis Generation: Cluster analysis is useful when a researcher wishes to develop hypothesis concerning the nature of the data or to examine previously stated hypotheses.
Objectives of Cluster Analysis

The primary objective of cluster analysis is to define the structure of the data by placing the most similar observation into groups. Cluster analysis is used for: taxonomy description, data simplification and relationship identification. Theoretical, conceptual and practical considerations must be observed when selecting clustering variables for cluster analysis:

1. Only variables that relate specifically to objectives of the cluster analysis are included; irrelevant variables cannot be excluded from the analysis once it begins.
2. Variables selected characterize the individuals (objects) being clustered.

Research Design in Cluster Analysis

With the objectives defined and variables selected, we must address four questions before starting the partitioning process.

1. Sample Size:
   a) Sufficient size is needed to ensure representativeness of the population and its underlying structure, particularly small groups within the population.
   b) Minimum group sizes are based on the relevance of each group to the research question and the confidence needed in characterizing that group.

2. Similarity be Measured:
   Similarities measures calculated across the entire set of clustering variables allow for the grouping of observations and their comparison to each other.
   a) Distance measures are most often used as a measure of similarity, with higher values representing greater dissimilarity (distance between cases) not similarity. Euclidean (straight line) distance, the most common measure of distance. Squared Euclidean distance, the sum of squared distances and the recommended measure for the centroids and ward’s method of clustering. Mahalanobis distance account for variable correlation and weights each variable equally; most appropriate when variables are highly correlated. Less frequently used are correlational measures, where large values do indicate similarity.

3. Outlier Detected:
   Outliers can severely distort the representativeness of the results if they appear as structure (clusters) inconsistent with the research objectives.

4. Data be Standardized:
   Clustering variables should be standardized whenever possible to avoid problems resulting from the use of different scale values among clustering variables. The most common standardization conversion is Z scores. If groups are to be identified according to an individual’s response style, then within-case or row-centering standardization is appropriate.

Assumptions in Cluster Analysis

Cluster analysis, like multidimensional scaling is not a statistical inference technique in which parameters from a sample or assessed as possibly being representative of a population. Instead, cluster analysis is a method for quantifying the structural characteristics of a set of observations. As such, it has strong mathematical properties but not statistical foundations. The requirement of normality, linearity and homoscedasticity that were so important in other techniques really has little bearing on cluster analysis. We must focus, however, on two other critical issues: representativeness of the sample, and
multicollinearity among variables in the cluster variate. Input variables should be examined for substantial multicollinearity.

**Deriving Clusters and Assessing Overall Fit**

With the clustering variables selected and the similarity matrix calculated, the partitioning process begins. The researcher must; Select the partitioning procedure used for forming clusters. Make the decision on the number of clusters to be formed.

**Two-Step Cluster Analysis:**

The Two-step Cluster Analysis procedure is an exploratory tool designed to reveal natural groupings (or clusters) within a data set that would otherwise not be apparent. The algorithm employed by this procedure has several desirable features that differentiate it from traditional clustering techniques:

1. The ability to create clusters based on both categorical and continuous variables.
2. Automatic selection of the number of clusters.
3. The ability to analyze large data files efficiently.
4. The two steps of the Two Step Cluster Analysis procedure's algorithm can be summarized as follows:

**5. RESULT AND DISCUSSION**

The results of Two Step Cluster Analysis technique are given in Appendix and discussion on more relevant results is as under.

Table 1 shows that frequency of each cluster. Total cases are 75192. Of the 49530 cases assigned to clusters, 40718 are assigned to the first cluster, 8816 to the second cluster.

Centroids of each cluster are shown in Table 2. Centroids are mean values of the variables. In cluster-1 we see that the mean of all variables are high as compares to second clusters, so we can say that non poor household is in first cluster and poor household in second cluster. In cluster one means non poor household shows average household income is 16510.7405, while in second cluster average household income is 4814.5938 it depicts that average household income is greater than in cluster one as compare to cluster two. On the average family size in non poor household is 6.7 while in second cluster which is of poor household has on the average family size 8.22. In poor household family size is greater than as compare to non poor household. Similarly on the average dependency ratio, sex ratio and education level of the household is greater than as compare to poor household.

The cluster frequency table is shown in Table 3 by poverty. Cluster 1 shows that those types of the household whose poverty level is non poor, with frequencies of 40714 as non poor household. Whereas cluster 2 shows those types of household whose poverty level is poor with frequency of 8816.

Figure 1 Shows that 95 percent confidence interval for centroids of household income with respect to their cluster. It contains the both cluster. The first cluster contain positive interval with upper limit is 16754.18 and lower control limit is 16267.30. In second cluster also contain the positive interval with upper limit is 4869.24 and lower limit is 4759.95. There is more variation in interval of household income with respect to cluster
one as compare cluster two. So household income in non poor household is high as compare to poor household.

Figure 2 Shows 95 percent confidence interval for centroids of family size with respect to their cluster. It contains the both cluster. The first cluster contain positive interval with upper limit is 6.7292 and lower control limit is 6.6645. It is represent significant interval. In second cluster contain the upper limit is 8.29 and lower limit is 8.14. Family size in poor household is higher as compare in non poor household.

Figure 3 contains positive interval in cluster 1 with upper limit is 0.3902 and lower limit is 0.3852. In second cluster the upper limit is 0.5176 and lower limit is 0.5085. The variation in interval one is small as compare to second interval. It means that in cluster of non poor household number of dependents are less as compare to the cluster of poor household. Due to large number of dependents low household income and large family size cause the household into poverty trap.

Figure 4 shows positive interval both cluster. In cluster 1 contains upper limit is 1.4154 and lower limit is 1.3927. In cluster 2 the interval shows more variation as compare to first cluster with upper limit is 1.2937 and lower control limit is 1.248. It is also significant.

Figure 5 shows the confidence interval for centroids of household education with respect to their cluster. It contains positive interval in cluster 1 with upper limit is 3.0561 and lower limit is 3.0231. In cluster 2 also contains positive interval with upper limit is 2.1343 and lower limit is 2.0758. This interval is also significant.

The "by variable" importance charts are produced with a separate chart for each cluster. The Y axis represents the variables, in descending order of importance. The dashed vertical lines represent the critical values for identified the significance of each variable. A variable is said to be significant when the value of t statistics exceeds the dashed line either positive or negative direction. A negative t statistic point out that the variable generally takes smaller than average values within this cluster, while a positive t statistic describe that the variable takes larger than average values. If the importance measures for all of the variables exceed the critical value in the chart, we can conclude that all of the continuous variables contribute to the formation of the cluster.

In figure 6, the chart for Cluster 1 shows that the factors (sex ratio, household income and household education) take larger than average values. These variables play significant role in formation of cluster one. The highest bar in cluster 1 is household education that supported in the construction of cluster 1, which means that this is the most significant factor in formation of non poor household cluster. The second important factor is household income take larger than average value in formation of cluster one. Dependency ratio and family size also play significant role in formation of cluster one its takes the smaller value then average value.

In figure 7, the chart for Cluster 2 the factors (dependency ratio and family size) take larger than average values while the factor household income, household education and sex ratio) takes smaller than the average values in formation of cluster 2.
6. CONCLUSION:

On this study we want to explore the factors which are playing a significant role for the cluster of non poor household and poor household and also investigate the direction and significance of importance variables. Two steps cluster analysis are used for data analysis purpose. Result demonstrated that (sex ratio, household income and household education) play positive role in non poor household. While dependency ratio and family size play negative role. But in poor household dependency ratio and family size play positive role and household income, household education and sex ratio play negative role. So that poverty increasing in poor household due to increasing family size, number of dependents and low household income. We suggest that Government should take step to increase the employment opportunities to young people to improve the standard of living by reducing poverty.

REFERENCES

### Table 1: Cluster Distribution

<table>
<thead>
<tr>
<th>Cluster Distribution</th>
<th>N</th>
<th>% of Combined</th>
<th>% of Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cluster</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>40714</td>
<td>82.2%</td>
<td>54.1%</td>
</tr>
<tr>
<td>2</td>
<td>8816</td>
<td>17.8%</td>
<td>11.7%</td>
</tr>
<tr>
<td>Combined</td>
<td>49530</td>
<td>100.0%</td>
<td>65.9%</td>
</tr>
<tr>
<td>Excluded Cases</td>
<td>25662</td>
<td></td>
<td>34.1%</td>
</tr>
<tr>
<td>Total</td>
<td>75192</td>
<td></td>
<td>100.0%</td>
</tr>
</tbody>
</table>

### Table 2: Cluster Centroids

<table>
<thead>
<tr>
<th>Centroids</th>
<th>Cluster 1</th>
<th>Cluster 2</th>
<th>Combined</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>Mean</td>
<td>Mean</td>
</tr>
<tr>
<td>household income</td>
<td>16510.7405</td>
<td>4814.5938</td>
<td>14428.9067</td>
</tr>
<tr>
<td>Family size</td>
<td>6.70</td>
<td>8.22</td>
<td>6.97</td>
</tr>
<tr>
<td>Dependency ratio</td>
<td>.3877</td>
<td>.5130</td>
<td>.4100</td>
</tr>
<tr>
<td>Sex ratio</td>
<td>1.4041</td>
<td>1.2709</td>
<td>1.3804</td>
</tr>
<tr>
<td>Household education</td>
<td>3.0396</td>
<td>2.1051</td>
<td>2.8733</td>
</tr>
</tbody>
</table>

### Table 3: Poverty

<table>
<thead>
<tr>
<th>Poverty</th>
<th>Poor household</th>
<th>Non poor household</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Frequency</td>
<td>Percent</td>
</tr>
<tr>
<td>Cluster</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>.0%</td>
</tr>
<tr>
<td>2</td>
<td>8816</td>
<td>100.0%</td>
</tr>
<tr>
<td>Combined</td>
<td>8816</td>
<td>100.0%</td>
</tr>
</tbody>
</table>
Fig. 1: Simultaneous 95% C.I for means

Fig. 2: Simultaneous 95% C.I for mean
Fig. 3: Simultaneous 95% C.I for means

Fig. 4: Simultaneous 95% C.I for mean
Classification of households with respect to poverty by using cluster analysis

**Fig. 5: Simultaneous 95% C.I for mean**

**Fig. 6: Continuous variables importance**
Fig. 7: Continuous variables importance
DETERMINANTS OF LABOUR PARTICIPATION:
EVIDENCE FROM PAKISTAN LABOUR FORCE SURVEY 2009-10

Zahoor Ahmad and Khadija Fatima
Department of Statistics, University of Gujrat, Gujrat, Pakistan
Email: zahoor.ahmed@uog.edu.pk; ufatima43@yahoo.com

ABSTRACT

The labour force participation is nearly universal for prime-aged persons in less developed countries, whereas its incidence is considerably lower for younger and older persons [Bloom and Freeman (1987)]. The objective of this paper is to investigate the role of socioeconomic and demographic determinants in participation of respondents in labour force. The data was obtained from Pakistan Labour Force Survey 2009-10 which was conducted by Federal Bureau of Statistics. The dependent variable has dichotomous response; participation and non participation and independent variables are demographic; age, sex, marital status, area and in socioeconomic variables used education levels of person’s, location of work and migration status. The Hierarchical Log-linear model has been used to identify model that best describing the relationship between these categorical variables. The results showed that factors such as age, sex, marital status, education level, and migration status and work location play a significant role in labour force participation.

KEYWORD

Participation; Socio-economic and Demographic Indicators; Pakistan Labour Force; Hierarchical Log-linear Model.

1. INTRODUCTION

Labor force participation means the proportion of the population who has ages 15 and above that is economically active population i.e. all people who supply labour force for the production of goods and services purpose during a specified period of time. Important aspects of development of the labor force are its growth in relation to the growth of population, capital, and the advance of technology; its composition in terms of sex, age groups, and other characteristics. The formation of quantities and qualities of the labor force cannot be independent of these factors, which are linked with the productivity and dynamism of the economy.

The labour force participation is nearly universal for prime-aged persons in less developed countries, whereas its incidence is considerably lower for younger and older persons [Bloom and Freeman (1987)]. Moreover, as a basis for evaluation of labor force participation between males and females in developed and developing countries, Bloom and Freeman (1987) found that the participation rates high for men than for the participation of women in both countries. However, Bloom and Freeman also found that,
the labour force participation rates for men have been falling while those for women have been increasing dramatically since the late of 1950s.

The youth in the Pakistan labour force facing numbers of barriers in their way of working, such like they early start to work, failing to enter in the labour market and also facing difficulties in moving across the jobs. Moreover, not having good education, lack of skills or training and having no experience are these major issues of youth in the labour market. In general, with increasing of age of persons labour force participation rate increases and unemployment decreases. Education plays a vital role in defining the youth activities. Young persons who are migrated from rural to urban earned two times more likely to be employed as compared to a person who does not migrate. The person who has technical/vocational training also affects the participation and employment in the labour market [Ahmad and Azim (2010)].

2. OBJECTIVE OF STUDY

The objective of this paper is to investigate the role of socioeconomics and demographic determinants in participation of respondents in labour force.

3. LITERATURE REVIEW

Salem, et al. (2008) carried out a study to explore the link between youth labour market and older workers labour force participation in the case of France. The results showed that the evidence of the correlation between youth labor market outcomes and old worker labor force participation plead more in favor of a positive association between young and old workers in the labor market. An increase in the old workers participation is indeed correlated with an increase in the employment rate of young workers and a decrease in their unemployment rate.

Ahmed and Azim (2010) conducted a study to analyze the youth labour market participation in Pakistan. In this study they used micro level data which was obtained from Labour Force Survey Pakistan 2006-07. This data set was collected from 32744 household based on about 224,280 individual about their age, sex, marital status, literacy, level of education, migration and other variables related to Employment and current activity of household members. The result showed that age, sex, marital status, migration, training, location, education level and characteristics of household have significant impact on employment probabilities of youth in Pakistan.

Balleer and Salvador (2011) developed a cohort based model of labour force participation to analyzed determinants of participation for disaggregated groups of workers in European countries, with a focus on the euro area. They used the age and cohort effects and observed determinants to construct trend measures and projections of labour supply in the euro area and the five largest euro area countries (Germany, France, Italy, Spain and the Netherlands). The results suggested that a decomposition based on age and cohort effects can account for a substantial part of the recent increase in labour force participation rates in the euro area, although not the surge since the mid-1990s. Cohort effects appear particularly relevant for women, with those born in the 1920s and 1930s less likely and those born in the late 1960s and early 1970s more likely to participate in the labour market over the life-cycle.
Congregado, et al. (2011) provided the empirical evidence on these two opposite effects in order to understand at least partially the why and wherefore of the recent evaluation of the Spanish labour force participation rate while attempting to anticipate its future evaluation. The estimates of their study reflect exactly the recent observed evolution in the Spanish participation rates and suggested that female participation rate will be still higher in the future whereas male participation rate has suffered a change in trend. In sum, the results suggest that from now on an equilibrium-distorting shock will have two opposite effects on the participation rate dynamic behavior depending on the gender.

4. DATA SOURCE AND ANALYSIS TECHNIQUES

The data for this study is taken from the Labour Force Survey (LFS) 2009-10 which was conducted by Federal Bureau of Statistics (FBS), Pakistan. The sample of 36400 households consists of 263501 individuals is given in the said survey. There are some missing cases in Labour Force Survey data related to variables age, sex, marital status, education levels, and location of work and migration status. So after excluding missing cases for any variable the ultimately sample size is 187732. We used Hierarchical Log-linear analysis for data analysis purpose.

5. RESULTS AND DISCUSSION

The log-linear analysis for model selection procedure is used to identify model that best describing the relationship between categorical variables and it analyzed in multi-way cross tabulations. Here we used to build the model by using backward elimination method and apply with saturated model with parameters estimates and test the partial association in terms of main effects, second order, third order and higher order. We have aim to investigate the association of these variables and also the strength of association by estimating the parameters values and odd ratio.

In this paper we want to investigate the role of such factors demographic; age, sex, marital status and socioeconomics; education levels, migration and location of work in participation of respondents in labour force. The valid sample size or valid cases are 187732 and weighted valid according to given factors are 81089 cases that contains all information on these variable.

The variables description and their categories level presented in table-1 age has form four categories in ordinal scale (10-35, 35-55, 56-75, >76), sex has two categories in nominal (1=male, 2=female), marital status has four categories in nominal scale (1= never marriage, 2= marriage, 3= widow, 4= divorce), education level has five categories with ordinal scale (No formal education, below matric, matric less inter, inter less degree, degree and above), migration status has also two categories with nominal scale (1= migrants, 2= non migrants), location of work has two categories ( 1 = urban, 2= rural) and participation has two categories with nominal scale (0 = non participation, 1= participation).

In table-2 contains information on k way and higher order effects and shows the probability that provides the observed significance level for the test that k way and higher
order effects are zero. Second part of table shows just k way effects it calculates by using upper portion of table and it also test that k way effects are zero. The likelihood chi-square with no parameters and have only the mean is 470598.942 and first order effect is 59958.961 shown in upper part of table-2. The difference between mean and first order effect is 470598.942 - 59958.961 = 410639.981 which displayed on the first line of the second portion of this table. This difference is measure of knowing about how much the model improve when first order effects are included. The small sig. (.0000) value shows that the hypothesis of first order effects being zero is rejected and accept alternative hypothesis which means that first order effect is present.

Now the difference is calculated as same procedure applies to the second order effect, the difference 59958.961 - 2107.438= 57851.523 displayed on the second line of second portion the same table-2. And the addition of second order effect improves the likelihood chi-square by 57851.523 this is also significant it means that there is second order effect is present in model. The difference between 2107.438- 319.322= 1788.116 is display on the third line of second portion of table-2 and also significant third order effect and remaining forth order and higher order does not help because p-value is not significant.

Goodness of fit test statistics shown in table-3 it represents the value of Likelihood Ratio Chi- Square and Pearson Chi-Square with its d.f and sig. values. The small value of both chi-square statistics and the greater p value from significance value (.05) indicate a good model. According to given values in table-3 provides that our final model after backward elimination is perfect fit for the given data.

The partial association given in table-4 which contains the partial chi-square with its sig. in term of the main effects, first order interaction term, second order, third order, forth order, fifth order and sixth order interaction. This table shows the significance of partial association in all these terms and in this table the partial associations are significant up to 3rd order interaction terms and remaining forth and higher order interactions are insignificant this reflects the decision of k-way and higher order effects in table-2.

Table-4 shows that all main effects age, sex, marital status, educational level, location of work, training and migration status are significant this indicate that all these variables are significantly related with participation of respondents and also mostly combinations of second order interaction parameters have significant partial association like age and sex, age and marital status, sex and marital status, age and location of work, sex and location of work, marital status and work location, age and educational level, sex educational level, marital status and education level, age and migration status, marital status and migration, educational level and migration, location of work and migration.

According to table-4 also mostly third order interactions terms are significant like Age, Sex and Marital status, Age, Sex and Work Location, Sex, Marital status and Work Location, Age, Sex and educational level, Age Marital Status and Educational level, Sex, Marital Status and Educational level, Age, Work Location and educational level, Sex, work location and Educational level, Marital Status, Work Location and Educational level, Sex, Marital status and migration, Marital Status, Educational level and Migration so on shown in table-4 and all others higher order interaction terms are not significant.
The parameter estimated value for main effects, odd ratios and 95% confidence interval for odd ratio is presented in table-5. The positive value of estimate of parameters shows the positive relationship and increase in response category and negative value show negative relationship and decrease in response category. The odd ratio tells us both direction of relationship and also provides the magnitude or strength of association. In this table we calculated odd ratio by taking the antilogarithm of their estimated parameters. Age have four categories but the forth category deal as a reference category for others and provide three estimates for age variable and age provide significant effect for three categories. In case of sex variable has two categories and first category of male and 2nd is female and compare male to female so sex has significant effect in participation. Marital status has four categories and forth category used as a reference category and never married vs. divorce is not significant but other two categories have significant effect. Migration has positively significant impact on participation of respondents in labour force. Location of work has negative effect in participation of respondents. The educational level has five categories and last category take as a reference category all others compare with it first two has positive impact on participation and other two have negative effect.

Table-6 represents the estimated parameter values of interaction effects, Odd Ratio and 95% Confidence Interval for Odd Ratio. Age has significant effect for three categories with participation rate. If we compare age group of 10-35 years of age with >76 then participation of respondents are in 10-35 years age of persons less than >76 of age. According to odd ratio we interpret as the odds ratio is 0.523 times less (95% C-I 0.468 to 0.586) for those who have age is between 10-35 years than those who have age greater than 76 years. And for the second category of age group odd ratio indicates that, the odd ratio is 0.668 time less (95% C-I 0.598 to 0.745) for those who have age is between 35-55 years than those who have age greater than 76 years and for the last odd ratio of age category indicates, the estimated odd of participation of respondent for group of age 56-75 years are 1.217 times higher (95% C-I 1.082 to 1.368) than the age greater than 76 years. So these results shows that a mostly participation rate high at the age of 56 to 75 than all other three categories.

Interaction terms for sex and participation has significant effect and we interpret as the estimated odd of participation of respondent in labour force for male 0.105 times (95% C-I 0.729 to 0.835) the estimated odd for those who are female’s participant. Marital status and participation has significant effect for 2nd and 3rd category and not for 1st categories we interpret as the category marriage vs divorce has negative effect in participation and widow vs divorce has positive effect in participation. Migration has negatively significant effect in participation of respondents in labour force. The interaction term for location of work and participation is insignificant (see table-6).

The education level has significant effect for four categories in participation and interpret in terms of odd ratio as the interaction terms for 1st category for no formal education with participation have significant effect and interpret as the estimated odd for those who have no formal education having participation is 0.595 times (95% C-I 1.790 to 2.270) the estimated odd for those who have higher degree, the estimated odd for those who have below matric education having participation is 0.855 times (95% C-I 1.790 to 2.270) the estimated odd for those who have higher degree, the odd ratio is 1.240 times
higher (95% CI 0.749 to 0.975) for those who have no formal education than those who have higher degree (see table-6).

Table-7 represents the 1st order interaction effects with participation. Age and sex, age and marital status, sex and marital status, marital status and educational levels, work location and education levels and age and migration have jointly significant effect in participation of respondents in labour force.

6. CONCLUSION:

The findings of this study showed that the socioeconomic and demographic factors such as age, education, marital status and gender play a significant role in participation of respondents in labour force. Furthermore the findings showed that age is an important determinant for participation rate and participation in labour force is high at older persons age 50-65 than for the younger one. Education levels play a significant role and according to this finding mostly participation rate high at the higher level of education after matric to higher level of education. Marital status and migration also has significant effect with participation of respondents in labour force. Gender also plays a significant role in participation in labour force and participation high for females than males.

REFERENCES

<table>
<thead>
<tr>
<th>Variables</th>
<th>No of categories</th>
<th>Names of categories</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>4</td>
<td>Age 1 (10-35), age 2 (35-55), age 3 (56-75), age 4 (&gt;76)</td>
</tr>
<tr>
<td>Sex</td>
<td>2</td>
<td>Sex 1 (1=male), sex 2 (2=female)</td>
</tr>
<tr>
<td>Marital status</td>
<td>4</td>
<td>Marital status (1= never marriage, 2= marriage, 3= widow, 4= divorce)</td>
</tr>
<tr>
<td>Migration status</td>
<td>2</td>
<td>Migration (1= migrants, 2= non migrants)</td>
</tr>
<tr>
<td>Location of work</td>
<td>2</td>
<td>Work location ( 1 = urban, 2= rural)</td>
</tr>
<tr>
<td>Educational level</td>
<td>5</td>
<td>No formal education, below matric, matric less inter, inter less degree, degree and above</td>
</tr>
<tr>
<td>Participation</td>
<td>2</td>
<td>Participation (0 = non participation, 1= participation)</td>
</tr>
</tbody>
</table>

**Table-2: K-Way and Higher-Order Effects**

<table>
<thead>
<tr>
<th>K-way and Higher Order Effects&lt;sup&gt;a&lt;/sup&gt;</th>
<th>K</th>
<th>df</th>
<th>Likelihood Ratio</th>
<th>Pearson</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Chi-Square</td>
<td>Sig.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Chi-Square</td>
<td>Sig.</td>
</tr>
<tr>
<td>K-way and Higher Order Effects&lt;sup&gt;a&lt;/sup&gt;</td>
<td>1</td>
<td>1279</td>
<td>470598.942</td>
<td>.000</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>1265</td>
<td>59958.961</td>
<td>.000</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>1186</td>
<td>2107.438</td>
<td>.000</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>954</td>
<td>319.322</td>
<td>1.000</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>571</td>
<td>71.907</td>
<td>1.000</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>213</td>
<td>9.778</td>
<td>1.000</td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>36</td>
<td>.000</td>
<td>1.000</td>
</tr>
</tbody>
</table>

K-way Effects<sup>b</sup>

<table>
<thead>
<tr>
<th>K</th>
<th>df</th>
<th>Likelihood Ratio</th>
<th>Pearson</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Chi-Square</td>
<td>Sig.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Chi-Square</td>
<td>Sig.</td>
</tr>
<tr>
<td>K-way Effects&lt;sup&gt;b&lt;/sup&gt;</td>
<td>1</td>
<td>14</td>
<td>410639.981</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>79</td>
<td>57851.523</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>232</td>
<td>1788.116</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>383</td>
<td>247.415</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>358</td>
<td>62.129</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>177</td>
<td>9.778</td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>36</td>
<td>.000</td>
</tr>
</tbody>
</table>

**Table-3: Goodness-of-Fit Test Statistics**

<table>
<thead>
<tr>
<th>Test Statistics</th>
<th>Test Statistics</th>
<th>df</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Likelihood Ratio Chi-Square</td>
<td>2.605</td>
<td>444</td>
<td>1.000</td>
</tr>
<tr>
<td>Pearson Chi-Square</td>
<td>1.643</td>
<td>444</td>
<td>1.000</td>
</tr>
<tr>
<td>Effect</td>
<td>df</td>
<td>Partial Chi-Square</td>
<td>p-value</td>
</tr>
<tr>
<td>------------------------------------</td>
<td>----</td>
<td>--------------------</td>
<td>---------</td>
</tr>
<tr>
<td>Age</td>
<td>3</td>
<td>74710.6</td>
<td>.000</td>
</tr>
<tr>
<td>Sex</td>
<td>1</td>
<td>31473.9</td>
<td>.000</td>
</tr>
<tr>
<td>Marital_Status</td>
<td>3</td>
<td>99569.2</td>
<td>.000</td>
</tr>
<tr>
<td>Work_Location</td>
<td>1</td>
<td>3378.3</td>
<td>.000</td>
</tr>
<tr>
<td>Educational_level</td>
<td>4</td>
<td>40898.9</td>
<td>.000</td>
</tr>
<tr>
<td>Migration</td>
<td>1</td>
<td>55295.9</td>
<td>.000</td>
</tr>
<tr>
<td>Participation</td>
<td>1</td>
<td>105314.2</td>
<td>.000</td>
</tr>
<tr>
<td>Age*Sex</td>
<td>3</td>
<td>1181.56</td>
<td>.000</td>
</tr>
<tr>
<td>Age*Marital_Status</td>
<td>9</td>
<td>30660.59</td>
<td>.000</td>
</tr>
<tr>
<td>Sex*Marital_Status</td>
<td>3</td>
<td>558.22</td>
<td>.000</td>
</tr>
<tr>
<td>Age*Work_Location</td>
<td>3</td>
<td>179.67</td>
<td>.000</td>
</tr>
<tr>
<td>Sex*Work_Location</td>
<td>1</td>
<td>1865.33</td>
<td>.000</td>
</tr>
<tr>
<td>Marital_Status*Work_Location</td>
<td>3</td>
<td>75.78</td>
<td>.000</td>
</tr>
<tr>
<td>Age*Educational_level</td>
<td>12</td>
<td>1446.92</td>
<td>.000</td>
</tr>
<tr>
<td>Sex*Educational_level</td>
<td>4</td>
<td>5140.16</td>
<td>.000</td>
</tr>
<tr>
<td>Marital_Status*Educational_level</td>
<td>12</td>
<td>425.421</td>
<td>.000</td>
</tr>
<tr>
<td>Work_Location*Educational_level</td>
<td>4</td>
<td>7321.02</td>
<td>.000</td>
</tr>
<tr>
<td>Age*Migration</td>
<td>3</td>
<td>550.99</td>
<td>.000</td>
</tr>
<tr>
<td>Sex*Migratino</td>
<td>1</td>
<td>194.66</td>
<td>.000</td>
</tr>
<tr>
<td>Marital_Status*Migration</td>
<td>3</td>
<td>262.01</td>
<td>.000</td>
</tr>
<tr>
<td>Work_Location*Migration</td>
<td>1</td>
<td>1558.14</td>
<td>.000</td>
</tr>
<tr>
<td>Educational_level*Migration</td>
<td>4</td>
<td>82.73</td>
<td>.000</td>
</tr>
<tr>
<td>Age<em>Sex</em>Marital_Status</td>
<td>9</td>
<td>32.494</td>
<td>.000</td>
</tr>
<tr>
<td>Age<em>Sex</em>Work_Location</td>
<td>3</td>
<td>25.459</td>
<td>.000</td>
</tr>
<tr>
<td>Age<em>Marital_Status</em>Work_Location</td>
<td>9</td>
<td>10.521</td>
<td>.310</td>
</tr>
<tr>
<td>Sex<em>Marital_Status</em>Work_Location</td>
<td>3</td>
<td>74.145</td>
<td>.000</td>
</tr>
<tr>
<td>Age<em>Sex</em>Edu_level</td>
<td>12</td>
<td>66.194</td>
<td>.000</td>
</tr>
<tr>
<td>Age<em>Marital_Status</em>Edu_level</td>
<td>36</td>
<td>74.917</td>
<td>.000</td>
</tr>
<tr>
<td>Sex<em>Marital_Status</em>Edu_level</td>
<td>12</td>
<td>385.886</td>
<td>.000</td>
</tr>
<tr>
<td>Age<em>Work_Location</em>Edu_level</td>
<td>12</td>
<td>79.567</td>
<td>.000</td>
</tr>
<tr>
<td>Sex<em>Work_Location</em>Edu_level</td>
<td>4</td>
<td>315.909</td>
<td>.000</td>
</tr>
<tr>
<td>Marital_Status<em>Work_Location</em>Edu_level</td>
<td>12</td>
<td>27.681</td>
<td>.006</td>
</tr>
<tr>
<td>Age<em>Sex</em>Migration</td>
<td>3</td>
<td>19.306</td>
<td>.000</td>
</tr>
<tr>
<td>Sex<em>Marital_Status</em>Migration</td>
<td>3</td>
<td>88.793</td>
<td>.000</td>
</tr>
<tr>
<td>Age<em>Work_Location</em>Migration</td>
<td>3</td>
<td>12.442</td>
<td>.006</td>
</tr>
<tr>
<td>Age<em>Sex</em>Migration</td>
<td>3</td>
<td>19.306</td>
<td>.000</td>
</tr>
<tr>
<td>Sex<em>Marital_Status</em>Migration</td>
<td>3</td>
<td>88.793</td>
<td>.000</td>
</tr>
<tr>
<td>Age<em>Work_Location</em>Migration</td>
<td>3</td>
<td>12.442</td>
<td>.006</td>
</tr>
<tr>
<td>Sex<em>Work_Location</em>Migration</td>
<td>1</td>
<td>11.876</td>
<td>.001</td>
</tr>
<tr>
<td>Marital_Status<em>Work_Location</em>Migration</td>
<td>3</td>
<td>32.162</td>
<td>.000</td>
</tr>
<tr>
<td>Age<em>Educational_level</em>Migration</td>
<td>12</td>
<td>21.743</td>
<td>.041</td>
</tr>
<tr>
<td>Sex<em>Educational_level</em>Migration</td>
<td>4</td>
<td>10.753</td>
<td>.029</td>
</tr>
<tr>
<td>Marital_Status<em>Educational_level</em>Migration</td>
<td>12</td>
<td>24.386</td>
<td>.018</td>
</tr>
<tr>
<td>Work_Location<em>Educational_level</em>Migration</td>
<td>4</td>
<td>59.443</td>
<td>.000</td>
</tr>
<tr>
<td>Marital_Status<em>Work_Location</em>Educational_level*Migration</td>
<td>12</td>
<td>22.679</td>
<td>.031</td>
</tr>
</tbody>
</table>
Table-5: Estimated Parameter values of main effects, Odd Ratio (OR) and 95% Confidence Interval for OR

<table>
<thead>
<tr>
<th>Variables</th>
<th>Estimated Parameters</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Age (years)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10-35 vs. &gt;76</td>
<td>.575</td>
<td>.000</td>
</tr>
<tr>
<td>36-55 vs. &gt;76</td>
<td>.284</td>
<td>.000</td>
</tr>
<tr>
<td>56-75 vs. &gt;76</td>
<td>-.193</td>
<td>.000</td>
</tr>
<tr>
<td><strong>Sex</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sex</td>
<td>.190</td>
<td>.000</td>
</tr>
<tr>
<td><strong>Migration</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Migrants vs. Non migrants</td>
<td>.173</td>
<td>.000</td>
</tr>
<tr>
<td><strong>Marital Status</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Never married vs. divorce</td>
<td>.024</td>
<td>.625</td>
</tr>
<tr>
<td>married vs. divorce</td>
<td>.815</td>
<td>.000</td>
</tr>
<tr>
<td>widow vs. divorce</td>
<td>-.241</td>
<td>.000</td>
</tr>
<tr>
<td><strong>Work Location</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Urban vs. Rural</td>
<td>-.212</td>
<td>.001</td>
</tr>
<tr>
<td><strong>Educational Level</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No Formal Education vs. Degree and Above</td>
<td>.354</td>
<td>.000</td>
</tr>
<tr>
<td>Below Matric vs. Degree and Above</td>
<td>.139</td>
<td>.014</td>
</tr>
<tr>
<td>Matric Less Inter vs. Degree and Above</td>
<td>-.105</td>
<td>.072</td>
</tr>
<tr>
<td>Inter But Less Degree vs. Degree and Above</td>
<td>-.241</td>
<td>.000</td>
</tr>
</tbody>
</table>

Table-6: Estimated Parameter values of interaction effects, Odd Ratio and 95% Confidence Interval for OR

<table>
<thead>
<tr>
<th>Variables</th>
<th>Estimated Parameters</th>
<th>Sig.</th>
<th>OR</th>
<th>95% CI for OR</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Age*Participation</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10-35 vs. &gt;76</td>
<td>-0.647</td>
<td>.000</td>
<td>0.523</td>
<td>0.468-0.586</td>
</tr>
<tr>
<td>36-55 vs. &gt;76</td>
<td>-0.404</td>
<td>.000</td>
<td>0.668</td>
<td>0.598-0.745</td>
</tr>
<tr>
<td>56-75 vs. &gt;76</td>
<td>0.196</td>
<td>.001</td>
<td>1.217</td>
<td>1.082-1.368</td>
</tr>
<tr>
<td><strong>Sex*Participation</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sex</td>
<td>-0.248</td>
<td>.000</td>
<td>0.105</td>
<td>0.729-0.835</td>
</tr>
<tr>
<td><em><em>Marital Status</em> Participation</em>*</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Never married vs. divorce</td>
<td>-0.006</td>
<td>.910</td>
<td>0.994</td>
<td>0.884-1.115</td>
</tr>
<tr>
<td>married vs. divorce</td>
<td>-0.972</td>
<td>.000</td>
<td>0.378</td>
<td>0.339-0.422</td>
</tr>
<tr>
<td>widow vs. divorce</td>
<td>0.253</td>
<td>.000</td>
<td>1.287</td>
<td>1.146-1.446</td>
</tr>
<tr>
<td><strong>Migration*Participation</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Migrants vs. Non migrants</td>
<td>-0.239</td>
<td>.000</td>
<td>0.787</td>
<td>0.736-0.842</td>
</tr>
<tr>
<td><strong>Work_Location*Participation</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Urban vs. Rural</td>
<td>-0.022</td>
<td>.510</td>
<td>0.978</td>
<td>0.915-0.989</td>
</tr>
<tr>
<td><em><em>Educational Level</em> Migration</em>*</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No formal Education vs. degree and above</td>
<td>-.519</td>
<td>.000</td>
<td>0.595</td>
<td>0.524-0.676</td>
</tr>
<tr>
<td>Below matric vs. degree and above</td>
<td>-.156</td>
<td>.020</td>
<td>0.855</td>
<td>0.749-0.975</td>
</tr>
<tr>
<td>Matric less inter vs. degree and above</td>
<td>0.159</td>
<td>.022</td>
<td>1.172</td>
<td>1.022-1.344</td>
</tr>
<tr>
<td>Inter but less degree vs. degree and above</td>
<td>0.334</td>
<td>.000</td>
<td>1.397</td>
<td>1.215-1.603</td>
</tr>
</tbody>
</table>
Table 7: Estimated Parameter values of 1st order interaction effects, Odd Ratio and 95% Confidence Interval for OR

<table>
<thead>
<tr>
<th>Variables</th>
<th>Estimated Parameters</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age<em>Sex</em>Participation</td>
<td>-0.128</td>
<td>.032</td>
</tr>
<tr>
<td>Age<em>Marital_Status</em>Participation</td>
<td>-1.09</td>
<td>.000</td>
</tr>
<tr>
<td>Sex<em>Marital_Status</em>Participation</td>
<td>-0.289</td>
<td>.000</td>
</tr>
<tr>
<td>Marital_Status<em>Educational_level</em>Participation</td>
<td>0.290</td>
<td>.012</td>
</tr>
<tr>
<td>Work_Location<em>Educational_level</em>Participation</td>
<td>0.214</td>
<td>.001</td>
</tr>
<tr>
<td>Age<em>Migration</em>Participation</td>
<td>-0.213</td>
<td>.000</td>
</tr>
</tbody>
</table>
CLASSIFICATION OF COUNTRIES WITH RESPECT TO THEIR GOVERNANCE

Zahoor Ahmad and Aysha Saleem
Department of Statistics, University of Gujrat, Gujrat, Pakistan
Email: zahoor.ahmed@uog.edu.pk; aysha_stat@gmail.com

ABSTRACT
Governance is a burning issue in the development literature because it almost affects every aspect of economy. If countries want better economic growth and human development then good reforms of governance is needed. Kaufmann, et al. (1999) developed six governance indicators named; Voice and Accountability, Political Stability, Government Effectiveness, Regulatory Quality, Rule of Law and Control of Corruption. In this paper the countries are classified with respect to their governance on the bases of governance indicators suggested by Kaufmann and those governance indicators are identified that play most important role for this classification. The findings reveal that Rule of Law and Control of Corruption are most significant indicators playing major role for the countries that are in good governance cluster and in bad Governance cluster respectively. Furthermore governance cannot be enhanced by considering just one dimension of it.

KEYWORDS
Governance; Governance Indicators; Two Step Clustering and Self Organizing Feature Maps.

1. INTRODUCTION

If countries want to increase the level of wellbeing of their people then they have to improve their economic growth and human development. Governance is one of the primary elements for the successful development in a country, especially good governance. That is for attaining the better economic growth and human development, in any economy, existence of good governance is crucial, especial in developing nations [Turner (2011)]. Governance usually plays a key role in the area like health, education, infrastructure, capital market regulation, economic stability, safety net provision, the legal system, creation of a good business environment, and the environment protection, etc, all of which are preconditions and basic features of the developed economy [(Brautigam 1991), (Landell-Mills and Seragedin 1991) and (Boeninger, 1992)]. The governance have significant effect on sustain the economic growth, development as well as human welfare, in a very long run. This is a massive effect that has been examined by many other authors [Kaufmann and Kraay (2002), Sebudubudu (2010), Pradhan and Sanyal (2011), Turner (2011)].

Governance is not a new concept. But in the recent past the term governance, specifically good governance, is being increasingly used in the development literature. Roughly speaking, the term governance cover all those aspects of the way a country are governed [Sharma (2007)]. Governance term is variously defined by different authors and organization. Some definitions of governance are given below.
United Nations Development Program (2002) describes the term governance as the “exercise of economic, political and administrative authority to manage a country’s affairs at all levels. It comprises the mechanisms, processes and institutions through which citizens and groups articulate their interests’ exercise their legal rights meet their obligations and mediate their differences”.

The World Bank economist Kaufmann, et al. (2009) describes the term governance as “The traditions and institutions by which authority in a country is exercised”. It has includes 1) The process, by which governments are selected, monitored and replaced. 2) The capacity of the government to effectively formulate and implement sound policies. 3) The respect of citizens and the state for the institutions that govern economic and social interactions. In sum, governance is a multidimensional concept which consists of political, economic and socio-cultural variables.

1.1 Good Governance:

The term good governance after its first introduction by the World Bank in 1989 has become increasingly popular and favorite among the donors. Good governance is now viewed as essential for promoting economic growth and alleviating poverty in the development countries by establishing conducive environment for saving and investment, risk taking, providing incentives to producers, and creating certainty in markets, increasing the size of markets by removing barriers to international trade and improvements in competitiveness [Haq and Zia (2009). Recently good governance has become conditionality for the disbursement of development assistance to less developed nations because fundamental donors and international financial organizations are increasingly basing their aid and loans on the condition that reforms that ensure good governance are guaranteed. Without good governance it is assumed that the benefits of the reforms will not reach to the poor and the funds will not be used effectively. [Santiso (2001), Sharma (2007)]. Additionally, foreign investors are also increasingly establishing the good governance, for their investment decisions, as a key to it [Fayissa and Nsiah (2010)].

As good governance is a wide concept it has been define differently by different organizations and others. Some definitions consider four dimensions of governance others six, yet others eight. Good governance has been defined as consisting of all or some combination of the following dimensions that are considered by Abdellatif (2003) and others: Participatory/Voice, Consensus Oriented, Accountable, Transparent, Responsive, Effective and Efficient, Equitable and Inclusive, Regulatory Quality, Political Stability and Rules of Law/Predictability. Currently many studies are conducted that represent the strong correlation between the good governance and economic performance having long term benefit. The authors also conclude in their studies that good governance is not only important to development but also that it is the leading factor in determining whether a country has the capability to use resources effectively to promote economic growth and reduce poverty [Ojima and Iimi (2005), Sharma (2007)].

1.2 Measuring Governance/Governance Indicators:

For the purposes of this study we are considering that definition of governance which is defined by the Kaufmann, et al. (1999b, 2002 and 2009). The Kaufmann, et al. (1999b, 2002 and 2009) defines some dimension of governance as follow.

i. Voice and Accountability

ii. Political Stability
iii. Government Effectiveness
iv. Regular Quality
v. Rule of Law
vi. Control of corruption

These indicators are used to measure the six dimensions of governance that are listed above. These dimensions can also be called the quality of governance. They cover more than 200 countries of the world. These six dimensions and indicators of governance are defined as follow by Kaufmann, et al. (1999b, 2002, 2009 and 2010) and Haq and Zia (2009).

The six indicators of governance are further divided into three parts that is first two indicators “Voice and Accountability” and “Political stability” used to measure the first perspective of definition that is the process, by which governments are selected, monitored and replaced. Next “Government Effectiveness” and “Regulatory Quality” are used to measure the second perspective of the definition which is the capacity of the government to effectively formulate and implement sound policies. Last two indicators “Rule of Law” and “Control of Corruption” measures the last perspective of the definition namely the respect of citizens and the state for the institutions that govern economic and social interactions among them. Some description is given below.

I) Political Governance:
The political governance group contains two indicators that are intended to capture the process by which government is selected, monitored and replaced. These two indicators are
i. Voice and Accountability (VA)
ii. Political Stability (PS)

First indicator ‘Voice and Accountability’ Measure the extent to which a country's citizens are able to participate in selecting their government, as well as freedom of expression, freedom of association, and a free media. It includes a number of indicators measuring various aspects of political process, civil liberties and political rights. While the ‘Political Stability’ indicator Measure the perceptions of the likelihood that the government will be destabilized or overthrown by unconstitutional or violent means, including politically-motivated violence and terrorism.

II) Economic Governance:
Two indicators are employed to summarize various indicators that include the government’s capacity to effectively formulate and implement sound policies. These are
i. Government Effectiveness (GE)
ii. Regulatory Quality (RQ)

The “government effectiveness” used to measure the quality of public services, the quality of the civil service and the degree of its independence from political pressures, the quality of policy formulation and implementation, and the credibility of the government's commitment to such policies. The “regulatory quality” governance indicator measures the ability of the government to formulate and implement sound policies and regulations that permit and promote private sector development.
III) Institutional Dimension of Governance:
The final dimensions of governance indicators are summarized in broad terms as the respect of citizen and the state of institutions that govern their interactions. This aspect of definition is measured by
i. Rule of Law (RL)
ii. Control of Corruption (CC)

Rule of law, summarizes several indicators that measure the extent to which agents have confidence in and abide by the rules of society, and in particular the quality of contract enforcement, property rights, the police, and the courts, as well as the likelihood of crime and violence. Control of corruption measure the extent to which public power is exercised for private gain, including both petty and grand forms of corruption, as well as "capture" of the state by elites and private interests. This aspect of corruption differs somewhat, ranging from the occurrence of additional payment to get things done, to assess the effect of corruption on business environment, to measure grand corruption in political arena or in the tendency of elites to engage in state capture. The presence of corruption is often a manifestation of a lack of respect on the part of both the corruptor and the corrupted for the rules that govern their interaction, thus represents a failure of governance.

1.3 Rationale and Significance:
In previous literature there is no such study exist that classify the countries with respect to their governance. Therefore we suggest that countries should classify with respect to their governance because previous studies showed that the countries that guarantee the good governance strategies are attractive places for the donors and foreign investors.

1.4 Objectives of Study:
In this study our basic purpose is to classify the countries with respect to their governance. Further those indicators of governance are determined that play most important role in classification.

2. LITERATURE REVIEW
Kaufmann, et al. (1999) represented some new governance indicators that are based on a variety of existing sources. These six aggregated governance indicators are capturing the six basic concepts of governance. PPP-adjusted per capita GDP, infant mortality and adult literacy are taken as development indicators. By utilizing regression it clarify that there is large and highly significant positive effects of governance on GDP. There is negative effect of good governance on the infant mortality and it also leads to a significant increase in the adult literacy. In the end it is concluded that governance matters for better development outcomes.

Yasuhsa and Atushi (2005) conducted a study to know that whether or not natural resource wealth and governance are interactively influence the economic development. Average growth rate of real GDP per capita is taken as dependent variable. Ordinary least square regression and cross country instrumental variable (IV) regression are used for data analysis. Results reveal that rich natural resources generally do not contribute to economic growth. Further the impact of government effectiveness has significant results interactively with resource abundance. These results shows that to generate the economic growth from the natural resources, it is required that government resource management should play a key role.
Ahmad and Aysha (2007) concentrated on whether good governance matters in development or not. It is mentioned that it is difficult to achieve rapid per capita income or improve the social indicators without improving the good governance indicators such that rule of law, voice and democratic accountability, stable political regimes, government effectiveness and control of corruption. Further it is concluded that strong accountable and effective political institutions are needed and a strong and long term commitment is required by aid agencies to provide funds and expertise to support governance reform project.

Haq and Zia (2009) conducted a study to examine the relation between the pro-poor economic growth and governance in Pakistan for 1996 to 2005 period. The six governance indicators (voice and accountability, political stability, government effectiveness, regulatory quality, rule of law and control of corruption) formulated by Kaufmann is taken. Poverty and income inequality is taken as the dimensions of the pro-poor governance. Empirical evidence is provided that the good governance in Pakistan can lead to the reduction in poverty and income inequality. Further it is suggested that Pakistan needs to implement sound and effective governance policies to achieve higher growth and millennium developmental goals.

Fayissa and Nsiah (2010) investigated the role of governance on economic growth (GDP per capita) for countries of the Sub-Saharan African region and investigate whether the impact of these governance indicators differs by conditional distribution of GDP per capita. The six governance indicators by Kaufmann are utilized. Quantile regression is used to fulfill the objective of the study. The result reveals the same relationships that are expected between per capita GDP and explanatory variables.

Turner (2011) plans a study and use the penal data to determine whether or not sustain growth is depending on the good governance in developing nations specifically in Sub-Saharan African countries. Annual percentage growth of GDP is taken as the dependent variable. A regression model is applied to check the significance of the independent variables. From the empirical results it is represented that government effectiveness index, corruption control index and general government expenditure are highly significant for the economic growth. Further it is concluded that for the economic development, in developing countries, stable, effective and governance without corruption is very essential.

3. SOURCE OF DATA AND ANALYSIS TECHNIQUES

The data on governance indicators is taken from the World Bank’s project of Worldwide Governance Indicators (WGI). As many authors uses these governance indicators to measure governance in any country [Kaufmann, et al. (1999) and (2002) Yasuhisa and Atsushi (2005), Uddin and Joya (2007), Haq and Zia (2009), Charron, et al. (2010), Fayissa and Nsiah (2010) and Turner (2011)]. Moreover Two Step Cluster Analysis and Self Organizing Feature Maps (SOM), a neural network technique for clustering, is utilized to fulfill the objectives.

4. RESULTS AND DISCUSSIONS

Self Organizing Feature Maps (SOM) is type of neural network. It is very useful for grouping the high dimensional data, to extract the natural grouping in the data. SOM can
be utilized to solve the problems, clustering and classification if output variable is provided to the network. If just output variable is provided to the network then it is only used to solve the clustering problem in which the label to each neuron is assigned by examining each observation in each neuron. Where as if output variable is provided then network utilize it to assign the name to each neuron automatically. So in our problem an output variable is provided to the network from the K-Means clustering membership. The results of the SOM is given in the Appendix-A.

Figure A-1 shows the iterative training graph during the training in progress. As during this training the training error is decreased and having value approximately less than 0.2. The training is performed in two phase. At first phase the value of the learning rate is set to 0.1 that will decrease to 0.02 and the neighborhood is set at 3 that also decrease to 1. In the second phase learning rate is decreased to 0.01 and neighborhood is set to zero. Model Summary table A-1 displays the information of overall model performance. The profile (SOFM 6:6-16:1) of the network exhibit that it is SOFM network having 6 input variable and one output variable, and two layers; input layer and output layer, having 6 and 16 units respectively. This table also reports training (1.00000), selection (0.961538) and test (0.961538) performance have the reasonable values that is close to 1 and training (0.164353), selection (0.224082) and test (0.223557) error also representing values of error close to zero. This report that overall network’s performance is good, because of small values of error and large values of performance is a good indication of it.

As the basic goal of SOM, is visualization and reduction of high-dimensional data into low-dimensional map, so topology map is an attractive feature where map plots the similarities of the data by grouping similar data items together. The topological map displays the output layer graphically in two dimensions. When a case is run, each unit displays its proximity to the case using a solid black square (a larger square indicating greater proximity), and the winning unit (the one who's center is nearest to the input case) is highlighted. From the figure A-2 (a & b) the topology map for case in Bad Governance (BG) and Good Governance (GG) is given. As it can be see easily that in figure A-2(a) first unit is blacker then the other unit then it is the winning neuron means the first country is in this unit. Where as in figure A-2(b) sixth unit is blacker then other so at this time it is the winning neuron. Further for each neuron in the topological map, number of cases in it and its label is given. For example first neuron has 9 countries and they all are related to the bad governance. As the label is assigned to the neuron on the bases of that “assigned the name bad governance if at least 80% countries in this neuron are related to the bad governance”. Further as it is mention above that the neurons that have blacker colour have the less activation value and at the same time it is winner neuron. From the topology, the neurons which have label GG is make one cluster and the neurons that have label BG is making second cluster.

Network illustration is another magnetic function of the SOM results which is given in figure A-3(a & b). This is the basic network working behind the scene. Here is an advantage of colour where light colour of a neuron represents the less activation. These are the same network for which the topology map is appeared in the figure A-2 (a & b).

As it is mention above that if output variable is provided to the network than Self Organizing Maps perform two task clustering and classification. So the classification table
is available on table A-2 which shows that from total of 213 countries 119 are assigned to the bad governance and 91 are assigned to the good governance and three observations are excluded due to the missing data in one or more variables. Out of 119 countries that are assigned to the bad governance cluster 115 are correctly classified from the remaining four countries are classified wrongly. Same from the 91 countries that all are correctly classified to the good governance cluster. So percentage of correct predication for bad governance is 96.6387% and for good governance 100%. Measures of classification in this table signify that classification is done in a good manner by the network.

To identify the most important variable that contributes more in this network sensitivity analysis is performed (table A-3). Ratio and rank for each variable is appears. Ranks to the variables are assigned on the bases of ratio, that variable which have highest value of ratio will receive the rank one and identify as the most contributive variable. Here rule of law has the large value of the ratio and its rank is one. So it is the most important variable. The next contributive variables are control of corruption, voice and accountability, government effectiveness, regulatory quality and political stability. So rule of law and control of corruption is the most important indicator in this network.

As the SOM tells only about the relative importance of indicators so to see that which indicator plays its important role in which cluster so two step cluster analysis is performed and its basic two figures (A-4 & A-5) are given. The importance of the variables can be seen from these figures. In these graphs the variables are lined up in descending order of importance for each cluster. To determine the significance of each variable the vertical lines in the dashed form represents the critical values. If the variable is significant, then it must cross the value of the t-statistic either a positive or negative direction.

Figure A-4 represents that all of the variables are crossed the critical value, the value of the t statistic is provided for each indicator. Thus it can be concluded that all of the indicators contribute to the formation of the first cluster. The Control of Corruption indicator is one that contributes more in the formation of first cluster having the value of test statistic -15.33. The next important variables are Government Effectiveness and Rule of Law having the value of test statistic -13.51 and -13.21. Regulatory Quality, Voice and Accountability and Political Stability are the following important indicators and the values of the test statistics are -10.48, -7.62 and -5.5. As all the variables are contributed in the negative direction so indicators obtain smaller than average values within this cluster. Thus the least contributive variable is the Political Stability in the formation of the cluster one. Further it can be said that the countries in cluster one have bad governance and Control of Corruption, Government Effectiveness, Rule of Law and Regulatory Quality is in the most terrible condition in these countries.

The importance of variable in cluster two is given in figure A-5 represents that all of the variables are significant because they crossed the critical value. The most contributive variable in the formation of second cluster is Rule of Law having the value of test statistic 20.1. After that Regulatory Quality and Government Effectiveness are more important variables and the value of test statistic is 18.89 and 18.3. The subsequent essential indicators are Control of Corruption, Political Stability and Voice and Accountability the values of the test statistics are 15, 14.77 and 12.3. As all the variables are contributed in the positive direction so indicators obtain larger than average values within this cluster. Thus the least contributive variable is the Voice and Accountability in the formation of
the second cluster. Further the countries that belong to the cluster two have good governance and Rule of Law, Regulatory Quality, Government Effectiveness, and Control of Corruption, is in the most magnificent condition in these countries.

To check the effect of two indicators simultaneously response surface is given. It is a surface plotted in three dimensions, indicating the response of one or more variable (or a neural network) as two input variables are adjusted with the others held constant. All possible response surfaces (figure A-6 to figure A-11) of the first four important indicators are given which are Rule of Law, Control of Corruption, Government Effectiveness and Regulatory Quality.

First response surface is for the most important indicators rule of Law and control of corruption in figure A-6. Here it can be seen that rule of law and control of corruption alone cannot contribute to improve the governance. So to boost up the governance both indicators should have simultaneous effect. Next response surfaces (figure A-7 to A-11) indicating the same pattern of indicators, with somewhat variation, to enhance the governance. So these represent that considering one dimension of governance alone is not able to improve the governance in any country

5. CONCLUSION

As the objective of the study is to categorize the countries with respect to their governance and to determine those governance indicators that plays most important role in the formation of the clusters. Consequently rule of law and control of corruption are the most important indicators to classify the countries. But control of corruption plays most important role for the countries in bad governance and rule of law is the most contributive indicator for countries of good governance.

So Control of Corruption, Government Effectiveness, Rule of Law and Regulatory Quality is in the most horrible condition for the countries in bad governance and Rule of Law, Regulatory Quality, Government Effectiveness, and Control of Corruption, is in the most magnificent condition for countries in good governance. These are same variables but having different ranking, it can be seen from figure A-4 and A-5.

This is the same result as observed from the self organizing feature maps and two step cluster analysis. So on that bases it can be said that rule of law and control of corruption are the more contributive indicators in formation of any governance. But on the other side response surface illustrate that governance cannot be good by just considering one dimension. Countries have to concentrate on all the indicators to batter their governance so that batter economic and developmental conditions can be obtained to achieve country’s objectives.

REFERENCES


Self Organizing Feature Maps (SOFM)

Table A-1:
Model Summary Report:

<table>
<thead>
<tr>
<th>Profile</th>
<th>SOFM 6:6-16:1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Train Performance</td>
<td>1.000000</td>
</tr>
<tr>
<td>Select Performance</td>
<td>0.961538</td>
</tr>
<tr>
<td>Test Performance</td>
<td>0.961538</td>
</tr>
<tr>
<td>Train Error</td>
<td>0.164353</td>
</tr>
<tr>
<td>Select Error</td>
<td>0.224082</td>
</tr>
<tr>
<td>Test Error</td>
<td>0.223557</td>
</tr>
<tr>
<td>Inputs</td>
<td>6</td>
</tr>
</tbody>
</table>

Figure A-1: iterative Training Graph
Figure A-2(a): Topology Map for case in BG

Figure A-2(b): Topology Map for case in GG
Classification of countries with respect to their governance

Table A-2: Classification:

<table>
<thead>
<tr>
<th></th>
<th>Type of BG</th>
<th>Type of GG</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total</td>
<td>119</td>
<td>91</td>
</tr>
<tr>
<td>Correct</td>
<td>115</td>
<td>91</td>
</tr>
<tr>
<td>Wrong</td>
<td>4</td>
<td>0</td>
</tr>
<tr>
<td>Unknown</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Correct (%)</td>
<td>96.6387</td>
<td>100</td>
</tr>
<tr>
<td>Wrong (%)</td>
<td>3.3613</td>
<td>0</td>
</tr>
<tr>
<td>Unknown (%)</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

BG  Bad Governance
GG  Good Governance

Table A-3: Sensitivity Analysis:

<table>
<thead>
<tr>
<th>Variables</th>
<th>Ratio</th>
<th>Rank</th>
</tr>
</thead>
<tbody>
<tr>
<td>Voice and Accountability</td>
<td>1.398957</td>
<td>5</td>
</tr>
<tr>
<td>Political Stability</td>
<td>1.138798</td>
<td>6</td>
</tr>
<tr>
<td>Government Effectiveness</td>
<td>1.454622</td>
<td>3</td>
</tr>
<tr>
<td>Regulatory Quality</td>
<td>1.338688</td>
<td>4</td>
</tr>
<tr>
<td>Rule of Law</td>
<td>1.520129</td>
<td>1</td>
</tr>
<tr>
<td>Control of Corruption</td>
<td>1.378836</td>
<td>2</td>
</tr>
</tbody>
</table>
Variable importance Graph: by Variable

Two Step Cluster Number = 1

Two Step Cluster Number = 2
Classification of countries with respect to their governance

Figure A-6

Figure A-7
Figure A-8

Figure A-9
Classification of countries with respect to their governance

Figure A-10

Figure A-11
PROBLEMS IN THE USE OF EDUCATIONAL TECHNOLOGY

Aamna Saleem Khan\textsuperscript{1} and Muhammad Qasim Rind\textsuperscript{2}

Preston University, Islamabad, Pakistan
Email: \textsuperscript{1}aamnasalim@yahoo.com; \textsuperscript{2}dr_rind@yahoo.com

ABSTRACT

The objective of the study was to identify the problems in the use of educational technology. A sample of two hundred students, twenty teachers and five heads of Secondary Schools in Wah Cantonment area were selected. The instrument consisted of 30 items in each questionnaire; three questionnaires were statistical prepared by the researcher. The data were analyzed by applying suitable statistical method. Major results of the study indicated that there is a close relationship between educational technology and the training of teachers to make teaching-learning process effective. It was observed that trained teachers have ability to use educational technology as well as ability to prepare instructional material for teaching. It was also felt that there were various problems in using educational technology effectively. It was due to non-availability of sufficient funds and resources, lack of teaching aids, improper and less use of available instructional materials and resources, untrained and non-motivated teachers, non-appreciation attitude from heads etc. In order to overcome the problems and the hurdles appearing in the use of educational technology, it was recommended that sufficient funds may be provided for efficient use of technology, and intensive training in educational technology may be arranged for the teachers.

KEY WORDS

Educational Technology; Schools; Teachers; Students; Instructional Material.

1. INTRODUCTION

ALLAH has created man superior to all living and non-living things. Man is superior to all things due to his wisdom and consciousness. Wisdom is the use of logical thinking to solve the problems of life that can be acquired through observation, personal experiences and education. Education is the right of every man and woman. ALLAH said in the Holy Quran “education is compulsory for every man and woman”. Holy Prophet Muhammad P.B.U.H said, get education from cradle to grave. No one can deny the importance of education as it is the social function that serves the society in many ways. The quality of instruction has been questioned from many decades. The quality of education is directly related to the quality of instruction in the classroom at the grassroots level (Bibi, 2005).

For the support of teaching-learning process, technology is an essential term, although it is widely used in administration and management of education. The educational technology is also concerned with process e.g. in delivering learning materials, facilitating communication and assessment and feedback.

Technology means systematic application of scientific or other organized knowledge to practical task. Therefore educational technology is based on theoretical knowledge from different disciplines and experiential knowledge from educational practice.
Technology is a part of education. There is deep relationship between education and technology. Without technology, education cannot be more effective. Education without technology is just like a body without soul and tea without sugar.

Technology is not only the “tool” for the development of science but also the “change” in the social process (Aggarwal, 1995).

Educational technology is the use of technology to improve education. It is systematic, iterative process for designing instruction or training used to improve performance.

Educational technology is the incorporation of Internet and other information technologies into the learning experience. This is the term that grows along with the advancement that is made in the fields. Learning and teaching options have drastically changed (Natalie, n.d).

Technology is a part of education. There is deep relationship between education and technology. Without technology, education cannot be more effective. Education without technology is just like a body without soul and tea without sugar.

Technology is not only the “tool” for the development of science but also the “change” in the social process (Aggarwal, 1995).

Educational technology concerns with the systematic use of modern methods for teaching and learning. It involves teachers in a variety of roles, some of which are traditional, some still emerging in this definition; special consideration is given to the adaptive role of the teacher. One purpose of studying educational technology is surely to help us to make the best use of capabilities of individual teachers. Educational technology is fundamentally aimed at improving the efficiency of the educational system by increasing the rate, depth, precision and value of the learning (Gillani, 2005).

This technological escalation has bestowed upon education proliferation of equipment and materials which can assist in the reorganization and redefinition of educational experiences. In the past, most teaching depended almost entirely on verbal communication between teacher and student or written communication to the student from printed materials. Although these communication channels continue to play important roles in the learning process, now students are learning facts, skills and attitudes from pictures, television, recorded words, programmed lessons and dramatic renovations. With the technological touch a simple schoolhouse turns into a systematized learning center (Aggarwal, 1995).

According to UNESCO (1978) the concept of educational technology has passed through three stages of development. By 1967, it was referred to as audio-visual aids. Till 1975, it was known as methods, materials and techniques. By 1978, it was termed as systems analysis. Thus managers of educational technology moved from technicians to specialists and then to groups. At the first stage, the objectives were restricted to technical and practical skills, at the second stage these were limited to optimization of the teaching-learning process through media and at the third stage new attitudes and approaches are emerged.
1.1 PRINCIPLES IN THE USE OF TEACHING AIDS

Teaching aids prove effective only when they suit the teaching objectives and unique characteristics of the special group of learners. Following points may be kept in view in this regard:

1) They should suit the age-level, grade-level and other characteristics of the learners.
2) They should have specific educational values besides being interesting and motivating.
3) They should be the true representatives of the real things.
4) They should help in the realization of desired learning objectives (Bibi, 2005).

1.2 IMPORTANCE OF EDUCATIONAL TECHNOLOGY

In addition vicarious experience can be gained from still pictures, films, filmstrips, resource persons, simulations, mockups, television, and the like. The more concrete and realistic the vicarious experience, the more nearly it approaches the learning effectiveness of the first level. Of course, unless the learner realizes that he is dealing with a substitute; his learning may not be comparable to that of real-life learning.

Audio-visual aids or devices or technological media or learning devices are added devices that help the teacher to clarify, establish, co-relate and co-ordinate accurate concepts, interpretations and appreciations and enable him to make learning more concrete, effective, interesting, inspirational, meaningful and vivid. They help in completing the triangular process of learning viz, motivation, clarification, stimulation. The aim of teaching with technological media is clearing the channel between the learner and the things that are worth learning. The basic assumption underlying audio-visual aids is that learning clear the understanding of the students from sense experience. The teacher must “show as well as tell”. Audio-visual aids provides significant gains in informational learning, retention and recall, thinking and reasoning, activity interest imagination, better assimilation and personal growth and development. The aids are the stimuli for learning ‘why’, ‘how’, ‘when’ and ‘where’. The hard to understand principles are usually made clear by the intelligent use of skillfully designed instructional aids.

The audio-visual aids are the best motivators. The students reduce verbalism. They help in giving clear concepts and thus help to bring accuracy in learning. Clear images are formed when we see, hear, touch, taste and smell as our experiences are direct, concrete and more or less permanent. Learning through the sense becomes the most natural and consequently the easiest. It is beyond doubt that the first-hand experience is the best type of educative experience. But it is neither practicable nor desirable to provide such experience to pupils. Substituted experiences may be provided under such conditions. There are many inaccessible objects and phenomena. For example, it is not possible for an average man to climb the Mount Everest. There are innumerable such things to which it is not possible to have direct access so, in all such cases, these aids help us.

“Mere chalk and talk” do not help. Audio-visual aids give variety and provide different tools in the hands of the teacher. When audio-visual aids are employed, there is great scope for children to move about, talk, laugh and comment upon. Under such an atmosphere the students work because they want to work and not because the teacher wants them to work. Many visual aids offer opportunities to students to handle and manipulate things. Audio-visual aids contribute to increased retentively as they stimulate response of the whole organism to the situation in which learning takes place The use of
audio-visual aids enables the teacher to follow the maxims of teaching like ‘concrete to abstract’, ‘known to unknown’ and ‘learning by doing etc.

The use of a variety of audio-visual aids helps in meeting the needs of different type of students. Use of audio-visual aids stir the imagination, thinking process and reasoning power of the students and calls for creativity, and inventiveness and other higher mental activities on the parts of students and thus helps the development of higher faculties among the students. Use of audio-visual aids helps in the learning of other concepts, principles and solving the real problems of life by making possible the appropriate positive transfer of learning and training received in the classroom. A balanced, rational and scientific use of these aids develops motivation, attracts the attention and interest of the students and provides a variety of creative outlets for the utilization of their tremendous energy and thus keeps them busy in the class work in this way, the overall classroom environment becomes conducive to creative discipline (Aggarwal, 1995).

Due to all these facts, it is desirable to identify the problems in the use of educational technology for effective learning.

2. RESEARCH METHODOLOGY

2.1 SAMPLE

The sample was consisting of five head masters and head mistress of Government Secondary Schools having both arts and science stream in Wah Cantt.

The researcher were selected five Government Secondary Schools of Wah Cantt on the basis of the qualification of teachers as Masters and Graduation in both arts and science stream.

The researcher selected twenty teachers of approximate equal qualification in both arts and science by simple random sample. Two teachers of M.A, M. Sc qualification and two teachers of B.A, B. Sc qualification were randomly selected for research.

200 students were selected from five Government Secondary Schools. One section from science and one from arts of 9th and 10th class respectively was randomly selected. Ten students, five gifted and five slow learners from each section were selected on the basis of their previous result.

2.2 RESEARCH INSTRUMENT

A questionnaire was used as a research instrument for the collection of data. Three questionnaires were used; each questionnaire contained 30 multiple-choice items.

Questionnaire I (Head of the Schools)
- 2 questions were related to the problems faced by the teachers in using educational technology.
- 2 questions were related to the perception of heads about teachers.

Questionnaire II (Secondary School Teachers)
- 1 question was related to the relevancy of teaching with the present society.
- 2 questions were related to the contribution of heads for the availability of audio-visual aids.
- 6 questions were related to the abilities of teachers.
- 2 questions were related to the problems in using instructional material.
• 1 question was related to the relevancy of instructional material with new era.
• 1 question was related to the reasons not to use audio-visual aids by the teachers.
• 1 question was related to the problems regarding educational facilities (availability and use)

**Questionnaire III (Students of Secondary Schools)**

• 6 questions were related to the efforts of teachers for the use of audio-visual aids.
• 2 questions were related to the accuracy of the use of audio-visual aids.

### 2.3 COLLECTION OF DATA

Data was collected for study through questionnaire. The researcher herself was prepared the questionnaire approved by the supervisor. The researcher herself visited the schools. The principal gave approval to collect the data and introduced the researcher to the teachers and the students of 9th and 10th class. The questionnaires were given to the heads, teachers and students, they were returned after completing in the required time.

### 2.4 ANALYSIS OF DATA

The data was analyzed by using Percentage.

<table>
<thead>
<tr>
<th>Sr #</th>
<th>Items/Statements</th>
<th>Agree</th>
<th>To some extent</th>
<th>Disagree</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>f</td>
<td>%</td>
<td>f</td>
<td>%</td>
</tr>
<tr>
<td>Heads</td>
<td>Teachers faced difficulties to use educational technology during their teaching.</td>
<td>4</td>
<td>80%</td>
<td>1</td>
<td>20%</td>
</tr>
<tr>
<td></td>
<td>School has sufficient money to buy equipment and audio-visual aids for making teaching learning effective.</td>
<td>0</td>
<td>0%</td>
<td>2</td>
<td>40%</td>
</tr>
<tr>
<td>Teachers</td>
<td>School has sufficient money to buy equipment and instructional material for making teaching learning effective.</td>
<td>1</td>
<td>5%</td>
<td>8</td>
<td>40%</td>
</tr>
<tr>
<td></td>
<td>The available facilities of audio-visual aids are sufficient for courses.</td>
<td>1</td>
<td>5%</td>
<td>5</td>
<td>25%</td>
</tr>
</tbody>
</table>

Table 1 shows that majority (80%) of heads agreed that teachers are unable to use educational technology for their teaching due to lack of availability of facilities in the classroom and insufficient funds in schools. Majority (60%) of heads and (55%) of teachers agreed that school has not enough money to purchase the teaching aids and the teachers face financial problems to use them in the classroom instruction. Majority (70%) teachers were of the view that although the teaching facilities are available for teaching but they do not fulfill the requirement of classrooms’ teaching.

The graphical representation is shown in figure I and figure II.
Problems in the Use of Educational Technology

Figure I: Problems faced by the teachers in using Educational Technology (Heads)

It shows that 80% schools have lack of availability of facilities and 60% schools have insufficient funds.

Figure II: Problems faced by the teachers in using Educational Technology (Teachers)

It shows that 55% schools have lack of funds and 70% available facilities do not fulfill teaching requirements.

Table 2: Perceptions of Heads about Teachers

<table>
<thead>
<tr>
<th>Sr #</th>
<th>Items/Statements</th>
<th>Agree</th>
<th>To some extent</th>
<th>Disagree</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>f</td>
<td>%</td>
<td>f</td>
<td>%</td>
</tr>
<tr>
<td>Heads</td>
<td>All the teachers of schools have full command on using</td>
<td>1</td>
<td>20%</td>
<td>2</td>
<td>40%</td>
</tr>
<tr>
<td></td>
<td>audio-visual aids.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Mostly science teachers use audio-visual aids for their</td>
<td>3</td>
<td>60%</td>
<td>2</td>
<td>40%</td>
</tr>
<tr>
<td></td>
<td>effective teaching.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 2 shows that some (40%) of heads were of the opinion that teachers do not have full command to use audio-visual aids for the effective learning of students and to some extent they have the ability to use them. Majority (60%) were of the view that mostly science teachers gave more importance as compared to arts teachers in using these aids to make their teaching more effective.

The graphical representation is shown in figure III.

![Graphical representation showing the perceptions of heads about teachers.](image)

Figure III: Perceptions of Heads about Teachers (Heads)

It shows that 40% heads said that teachers have lack of handling techniques and 60% heads said that only science teachers use these aids.

Table 3: Efforts/Contribution of Heads and Teachers for the better use of Audio-Visual Aids

<table>
<thead>
<tr>
<th>Sr #</th>
<th>Items/Statements</th>
<th>Agree</th>
<th>To some extent</th>
<th>Disagree</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>T</td>
<td>1. Heads contributes in the selection and utilization of audio-visual aids.</td>
<td>8</td>
<td>7</td>
<td>5</td>
<td>20</td>
</tr>
<tr>
<td></td>
<td>2. The headmasters appreciate teachers when they teach classes by audio-visual aids.</td>
<td>13</td>
<td>3</td>
<td>4</td>
<td>20</td>
</tr>
</tbody>
</table>

| S    | 1. Teachers use modern technology to make their teaching attractive.             | 123   | 39             | 38       | 200   |
|      | 2. Teachers use audio-visual aids for the clarification of concepts.            | 103   | 40             | 57       | 200   |
|      | 3. Teachers encourage the students to prepare audio-visual aids.                | 122   | 45             | 33       | 200   |
|      | 4. Teachers use audio-visual aids according to the demand of topic.             | 122   | 41             | 41       | 200   |
|      | 5. Teachers use the available material effectively to achieve the target objectives. | 98    | 84             | 18       | 200   |
Table 3 shows that about (40%) of teachers agreed that heads help in the selection and utilization of audio-visual aids. Majority (65%) of teachers said that they take appreciation when the teachers teach their classes with the help of audio-visual aids.

Mostly (57%) students said that teachers use audio-visual aids to make their teaching attractive, for clarification of concepts, to achieve objectives according to the demand of topic and even teachers encourage the students to prepare audio-visual aids.

The graphical representation is shown in figure IV and figure V.

Figure IV: Contribution of heads and teachers for use of a-v aids (Teachers)

It shows that 40% heads help in the selection and utilization of a-v aids and 65% heads appreciate teachers to use a-v aids.

Figure V: Contribution of heads and teachers for use of a-v aids (Students)

It shows that 57% students were of the view that teachers use a-v aids to make their teaching effective, to clarify the concepts, to achieve objectives and even teachers encourage them to prepare a-v aids.
Table 4: Relevancy of Instructional Material with present society, new era and grade and age level of students

<table>
<thead>
<tr>
<th>Sr #</th>
<th>Items/Statements</th>
<th>Agree</th>
<th>To some extent</th>
<th>Disagree</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>f</td>
<td>%</td>
<td>f</td>
<td>%</td>
</tr>
<tr>
<td>Teachers</td>
<td>1. Teaching is according to the demand of present society</td>
<td>7</td>
<td>35%</td>
<td>13</td>
<td>65%</td>
</tr>
<tr>
<td></td>
<td>2. The present state of instructional material in school matched with the needs</td>
<td>5</td>
<td>25%</td>
<td>2</td>
<td>10%</td>
</tr>
<tr>
<td></td>
<td>of 21st century</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Students</td>
<td>1. Technology is very essential to get progress in educational field.</td>
<td>190</td>
<td>95%</td>
<td>10</td>
<td>5%</td>
</tr>
<tr>
<td></td>
<td>2. Instructional material is according to the grade level and age level of students</td>
<td>125</td>
<td>63%</td>
<td>67</td>
<td>34%</td>
</tr>
</tbody>
</table>

Table 4 shows that majority (65%) of the teachers were of the view that teaching is not 100 percent according to the demand of present society. Majority (65%) of the teachers disagreed about the matching of instructional material with the need of 21st century.

Majority (95%) of the students said that technology is very essential for the progress of educational field. Majority (63%) of the students agreed that instructional material used by the teachers is according to grade and age level of students.

The graphical representation is shown in figure VI and figure VII.

![Figure VI](image)

Figure VI: Relevancy of instructional material with present society, new era & grade and age level of students (Teachers)

It shows that 65% teachers said that a-v aids are non-relevant with present society and new era.
Figure VII: Relevancy of instructional material with present society, new era & grade and age level of students (Students)

It shows that 95% students are of the view of importance of technology for progress and 63% said that a-v aids are relevant to grade and age level.

Table 5: Abilities of Teachers

<table>
<thead>
<tr>
<th>Sr #</th>
<th>Items/Statements</th>
<th>Agree</th>
<th>To some extent</th>
<th>Disagree</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Teachers prepare audio-visual aids by themselves.</td>
<td>3</td>
<td>15%</td>
<td>8</td>
<td>40%</td>
</tr>
<tr>
<td>2.</td>
<td>Teachers use audio-visual aids to clear the concepts in the classes by audio-visual aids.</td>
<td>4</td>
<td>20%</td>
<td>11</td>
<td>55%</td>
</tr>
<tr>
<td>3.</td>
<td>Teachers evaluate the students’ achievements after teaching with the help of audio-visual aids.</td>
<td>8</td>
<td>40%</td>
<td>11</td>
<td>55%</td>
</tr>
<tr>
<td>4.</td>
<td>Trained teachers can assess students, understanding easily by the use of effective instructional material.</td>
<td>15</td>
<td>75%</td>
<td>4</td>
<td>20%</td>
</tr>
<tr>
<td>5.</td>
<td>Competent teachers not only rely on the textbooks but he/she also consults other instructional material.</td>
<td>19</td>
<td>95%</td>
<td>00</td>
<td>0%</td>
</tr>
<tr>
<td>6.</td>
<td>Competent teachers try to provide current knowledge with the help of up to date instructional material.</td>
<td>14</td>
<td>70%</td>
<td>6</td>
<td>30%</td>
</tr>
</tbody>
</table>

Table 5 shows that about (45%) of the teachers said that teacher do not prepare audio-visual aids by themselves. Majority (55%) of the teachers said that to some extent teachers use audio-visual aids to clarify the concepts and to evaluate the students’ achievement. Majority (75%) of the teachers agreed that they assess students’ understanding easily by using audio-visual aids. Majority (95%) of the teachers agreed
that competent teachers not only rely on the textbook but also consult other instructional material. Majority (70%) of the teachers said that competent teachers try to provide current knowledge by using up-to-date instructional material.

The graphical representation is shown in figure VIII.

![Figure VIII: Abilities of Teachers (Teachers)](image)

It shows that 45% teachers do not prepare teaching aids, 55% teachers rarely use a-v aids for clarification of ideas and for students’ assessment, 75% teachers said that assessment is possible by a-v aids, 95% teachers said that only competent teachers use a-v aids in a better way and 70% teachers said that competent teachers provide current knowledge by using a-v aids.

| Table 6: Problems regarding Educational Facilities (Availability and Use) |
|---|---|---|---|---|
| Sr # | Items/Statements | Agree | To some extent | Disagree | Total |
|     |                | f % | f % | f % | f % | f % | % |
| Teachers |
| 1. | Medium of instruction is not according to the demand of present society. | 15 | 75% | 3 | 15% | 2 | 10% | 20 | 100% |
| 2. | Non serious attitude of Government for the improvement of education. | 6 | 40% | 7 | 35% | 7 | 35% | 20 | 100% |
| 3. | Lack of teaching aids and educational facilities. | 17 | 85% | 3 | 15% | 00 | 0% | 20 | 100% |
| 4. | Lack of special training of teachers. | 15 | 75% | 4 | 20% | 1 | 5% | 20 | 100% |

Table 6 shows that majority (75%) of the teachers said that medium of instruction is the problem of using educational facilities as it is not according to the demand of present society. About (40%) of the teachers agreed that educational facilities are not properly used due to non-serious attitude of government for the improvement of education. Majority (85%) of the teachers agreed that due to lack of teaching aids and educational facilities, teachers are not able to utilize them properly. Majority (75%) of the teachers were of the opinion that lack of proper training of teachers is another problem of using audio-visual aids for effective teaching.
The graphical representation is shown in figure IX.

![Graphical representation showing educational technology problems](image)

Figure IX: Problems regarding Educational Facilities (Availability & Use) (Teachers)

It shows that 69% teachers identify medium of instruction, non-serious attitude of Government, lack of teaching aids and proper training of teachers as problems in the use of a-v aids.

**Table 7: Reasons not to use the Audio-Visual Aids by the Teachers**

<table>
<thead>
<tr>
<th>Sr #</th>
<th>Items/Statements</th>
<th>Agree</th>
<th>To some extent</th>
<th>Disagree</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>f</td>
<td>%</td>
<td>f</td>
<td>%</td>
</tr>
<tr>
<td>Teachers 1.</td>
<td>The course is very lengthy</td>
<td>12</td>
<td>60%</td>
<td>8</td>
<td>40%</td>
</tr>
<tr>
<td>2.</td>
<td>They have to spend money from their own pocket.</td>
<td>8</td>
<td>40%</td>
<td>7</td>
<td>35%</td>
</tr>
<tr>
<td>3.</td>
<td>Audio-visual aids things are low use in their practical life.</td>
<td>4</td>
<td>20%</td>
<td>11</td>
<td>55%</td>
</tr>
<tr>
<td>4.</td>
<td>Audio-visual aids are not available in school.</td>
<td>1</td>
<td>5%</td>
<td>5</td>
<td>25%</td>
</tr>
</tbody>
</table>

Table 7 shows that majority (60%) of the teachers said that due to lengthy course they are not able to use audio-visual aids. About (40%) of the teachers said that for preparation of audio-visual aids, they have to spend money from their own pocket. They are expensive, so it is not possible to buy them for each topic or subject. Majority (55%) of the teachers said that to some extent, they are not relevant with practical life of students so they are reluctant to use them in classroom. Majority (70%) of the teachers disagreed with the statement that audio-visual aids are not available in school or classroom; they said that they are present but they do not use them due to many other reasons.

The graphical representation is shown in figure X.
Figure X: Reasons of not using a-v aids by teachers (Teachers)

It shows that 52% teacher said that due to lengthy course, use of their own money and non-relevancy of them with practical life, they are not able to use them properly and 70% teachers said that although a-v aids are present but they are not using them due to many other reasons.

3. DISCUSSION AND CONCLUSION

The research was aimed at to identify the problems in the use of instructional material for 10th class students of Wah Cantt Area studying in schools.

This study supports the other research studies i.e. Krysa, 1998 and Mumtaz, 2000. Their results showed that major factors that effects on the implementation of instructional material are limited accessibility of teachers and students, lack of technical and others, teachers support, lack of familiarity and teaching experiences, teachers’ attitude towards the use of audio-visual aids, personal knowledge, ease of use, incentive to change, support of schools etc. The present study indicated various problems for using educational technology in a class i.e. non-availability of sufficient funds and resources, lack of teaching aids, improper and less use of available instructional materials and resources, untrained and non-motivated teachers, non-appreciation attitude from heads, less participation of students, non-relevancy of teaching aids with practical life, society, new era and even age and grade level of students, lack of teachers’ abilities to prepare them, lengthy courses etc.

4. RECOMMENDATIONS

Keeping in the view the results of the study, the following recommendations are made:

1. To meet the modern challenges, Principals of schools should motivate teachers to use audio-visual aids in the classroom teaching.
2. Educational technology should be provided to schools for changing the students’ behaviour by providing the first hand knowledge to achieve the desirable objectives.
3. Financial resources should be allocated for audio-visual aids to the Principals of schools.
4. Teachers should motivate the students to search the related material about the topic to enhance their learning, for the development of understanding and for capturing their attention.
5. It is also suggested that the best quality of instructional material should be provided to schools so that the audio-visual aids can be useful for a longer period.
6. It is recommended that to overcome all the hurdles in the use of teaching aids by giving proper funds, training to the teachers and by reducing the length of chapters in the course.
7. Improvement may be brought in the quality of teaching by giving in-service training to the teachers according to the needs.

REFERENCES
COMPUTER HACKING FORENSICS INVESTIGATION TECHNIQUES

Nadia Qasim¹, Muhammad Qasim Rind² and Sheikh Muhammad Saleem²
¹ Kings College London, Email: nadia.nadya@gmail.com
² Preston University, Islamabad, Pakistan
   Email: dr_rind@yahoo.com; drsmsaleem@yahoo.com

ABSTRACT

Computer hacking forensic investigation is the process of detecting hacking attacks and properly extracting evidence to report the crime and conduct audits to prevent future attacks. Computer forensics is simply the application of computer investigation and analysis techniques in the interests of determining potential legal evidence. Evidence might be sought in a wide range of computer crime or misuse, including but not limited to theft of trade secrets, theft of or destruction of intellectual property, and fraud. Computer hacking forensic investigators can draw on an array of methods for discovering data that resides in a computer system, or recovering deleted, encrypted, or damaged file information.

KEYWORDS

Computer hacking; Computer investigation; Computer crime.

1. INTRODUCTION

Computer crime in today’s cyber world is on the rise. Computer Investigation techniques are being used by police, government and corporate entities globally. In the early 1980s personal computers became more accessible to consumers leading to their increased use in criminal activity (for example, to help commit fraud). At the same time, several new "computer crimes" were recognized (such as hacking). The discipline of computer forensics emerged during this time as a method to recover and investigate digital evidence for use in court. Today it is used to investigate a wide variety of crime, including child pornography, fraud, cyber stalking, murder and rape. The discipline also features in civil proceedings as a form of information gathering (for example, Electronic discovery).

Forensic techniques and expert knowledge are used to explain the current state of a digital artifact; such as a computer system, storage medium (e.g. hard disk or CD-ROM), an electronic document (e.g. an email message or JPEG image). The scope of a forensic analysis can vary from simple information retrieval to reconstructing a series of events. In a 2002 book Computer Forensics authors Kruse and Heiser define computer forensics as involving "the preservation, identification, extraction, documentation and interpretation of computer data". They go on to describe the discipline as "more of an art than a science", indicating that forensic methodology is backed by flexibility and extensive domain knowledge.
Thus, Computer forensics is a branch of digital forensic science pertaining to legal evidence found in computers and digital storage media. The goal of computer forensics is to examine digital media in a forensically sound manner with the aim of identifying, preserving, recovering, analyzing and presenting facts and opinions about the information.

Although it is most often associated with the investigation of a wide variety of computer crime, computer forensics may also be used in civil proceedings. The discipline involves similar techniques and principles to data recovery, but with additional guidelines and practices designed to create a legal audit trail. Computer forensic investigations usually follow the standard digital forensic process (acquisition, analysis and reporting). Investigations are performed on static data (i.e. acquired images) rather than "live" systems. This is a change from early forensic practices which, due to a lack of specialist tools, saw investigations commonly carried out on live data.

2. COMPUTER HACKING FORENSIC INVESTIGATION

Computer hacking forensic investigation is the process of detecting hacking attacks and properly extracting evidence to report the crime and conduct audits to prevent future attacks. Computer forensics is simply the application of computer investigation and analysis techniques in the interests of determining potential legal evidence. Evidence might be sought in a wide range of computer crime or misuse, including but not limited to theft of trade secrets, theft of or destruction of intellectual property, and fraud.

Evidence from computer forensics investigations is usually subjected to the same guidelines and practices of other digital evidence. It has been used in a number of high profile cases and is becoming widely accepted as reliable within US and European court systems.

Securing and analyzing electronic evidence is a central theme in an ever-increasing number of conflict situations and criminal cases. Electronic evidence is critical in the situations like disloyal employees, computer break-ins, possession of pornography, breach of contract, industrial espionage-mail fraud, bankruptcy, disputed dismissals, web page defacements, and theft of company documents.

Computer forensics enables the systematic and careful identification of evidence in computer related crime and abuse cases. This may range from tracing the tracks of a hacker through a client’s systems, to tracing the originator of defamatory emails, to recovering signs of fraud.

3. DIGITAL FORENSIC INVESTIGATION PROCESS

The Digital forensic process is a recognized scientific and forensic process used in digital forensics investigations. Forensics researcher Eoghan Casey defines it as a number of steps from the original incident alert through to reporting of findings. The process is predominantly used in computer and mobile forensic investigations and consists of three steps: acquisition, analysis and reporting.
3.1 Digital Media Seizure

Digital media seized for investigation is usually referred to as an "exhibit" in legal terminology. Investigators employ the scientific method to recover digital evidence to support or disprove a hypothesis, either for a court of law or in civil proceedings.

Prior to the actual examination digital media will be seized. In criminal cases this will often be performed by law enforcement personnel trained to as technicians so as to ensure preservation of evidence. In civil matters it will usually be a company officer, often untrained. Various laws cover the seizure of material. In criminal matters law related to search warrants is applicable. In civil proceedings the assumption is that a company is able to investigate their own equipment without a warrant, so long as the privacy and human rights of employees are observed.

3.2 Imaging or Acquisition Stage

Once exhibits have been seized an exact sector level duplicate (or "forensic duplicate") of the media is created, usually via a write blocking device, a process referred to as Imaging or Acquisition. The duplicate is created using a hard-drive duplicator or software imaging tools such as DCFLdd, IXimager, Guymager, TrueBack, EnCase, FTK Imager or FDAS. The original drive is then returned to secure storage to prevent tampering.

The acquired image is verified by using the SHA-1 or MD5 hash functions. At critical points throughout the analysis, the media is verified again, known as "hashing", to ensure that the evidence is still in its original state. In corporate environments seeking civil or internal charges, such steps are generally overlooked due to the time required to perform them.

3.3 Data Analysis Stage

After acquisition the contents of image files are analyzed to identify evidence that either supports or contradicts a hypothesis or for signs of tampering (to hide data). In 2002 the International Journal of Digital Evidence referred to this stage as "an in-depth systematic search of evidence related to the suspected crime". By contrast Brian Carrier, in 2006, describes a more "intuitive procedure" in which obvious evidence is first identified after which "exhaustive searches are conducted to start filling in the holes".

During the analysis an investigator usually recovers evidence material using a number of different methodologies (and tools), often beginning with recovery of deleted material. Examiners use specialist tools (EnCase, FTK, etc.) to aid with viewing and recovering data. The type of data recovered varies depending on the investigation; but examples include email, chat logs, images, internet history or documents. The data can be recovered from accessible disk space, deleted (unallocated) space or from within operating system cache files.

Various types of techniques are used to recover evidence, usually involving some form of keyword searching within the acquired image file; either to identify matches to relevant phrases or to parse out known file types. Certain files (such as graphic images) have a specific set of bytes which identify the start and end of a file, if identified a deleted file can be reconstructed. Many forensic tools use hash signatures to identify notable files or to exclude known (benign) ones; acquired data is hashed and compared to
pre-compiled lists such as the Reference Data Set (RDS) from the National Software Reference Library.

3.4 Reporting Stage

Once evidence is recovered the information is analyzed to reconstruct events or actions and to reach conclusions, work that can often be performed by less specialist staff. Digital investigators, particularly in criminal investigations, have to ensure that conclusions are based upon data and their own expert knowledge. Keeping in view following facts:

i) the testimony is based upon sufficient facts or data,
ii) the testimony is the product of reliable principles and methods,
iii) the witness has applied the principles and methods reliably to the facts of the case

4. COMPUTER FORENSICS INVESTIGATION TECHNIQUES

A number of techniques are used during computer forensics investigations some are listed below:

4.1 Cross-drive analysis

A forensic technique that correlates information found on multiple hard drives. The process, which is still being researched, can be used for identifying social networks and for performing anomaly detection.

4.2 Live analysis

The examination of computers from within the operating system using custom forensics or existing sysadmin tools to extract evidence. The practice is useful when dealing with Encrypting File Systems, for example, where the encryption keys may be collected and, in some instances, the logical hard drive volume may be imaged (known as a live acquisition) before the computer is shut down.

4.3 Deleted files

A common technique used in computer forensics is the recovery of deleted files. Modern forensic software has their own tools for recovering or carving out deleted data. Most operating systems and file systems do not always erase physical file data, allowing it to be reconstructed from the physical disk sectors. File carving involves searching for known file headers within the disk image and reconstructing deleted materials.

4.4 Volatile data

When seizing evidence, if the machine is still active, any information stored solely in RAM that is not recovered before powering down may be lost. One application of "live analysis" is to recover RAM data (for example, using Microsoft's COFEE tool, windd, WindowsSCOPE) prior to removing an exhibit.

RAM can be analyzed for prior content after power loss, because the electrical charge stored in the memory cells takes time to dissipate. The length of time for which data recovery is possible is increased by low temperatures and higher cell voltages. Holding unpowered RAM below −60 °C will help preserve the residual data by an order of magnitude, thus improving the chances of successful recovery. However, it can be impractical to do this during a field examination.
5. DIGITAL FORENSICS ANALYSIS TOOLS

A number of open source and commercial tools exist for computer forensics investigation. The details Forensic software tools for Windows may be seen at annexure-1. Typical forensic analysis includes a manual review of material on the media, reviewing the Windows registry for suspect information, discovering and cracking passwords, keyword searches for topics related to the crime, and extracting e-mail and pictures for review.

5.1 EnCase® Forensic software

EnCase® Forensic, the industry-standard computer investigation solution, is for forensic practitioners who need to conduct efficient, forensically sounds data collection and investigations using a repeatable and defensible process. The proven, powerful, and trusted EnCase® Forensic solution, lets examiners acquire data from a wide variety of devices, unearth potential evidence with disk level forensic analysis, and craft comprehensive reports on their findings, all while maintaining the integrity of their evidence.

EnCase® provides investigators with a single tool, capable of conducting large-scale and complex investigations from beginning to end. Law enforcement officers, government/corporate investigators and consultants around the world benefit from the power of EnCase® Forensic software. It acquires data in a forensically sound manner using software with an unparalleled record in courts worldwide. It investigates and analyzes multiple platforms such as Windows, Linux, AIX, OS X, Solaris etc using a single tool. It also save days, if not weeks, of analysis time by automating complex and routine tasks with prebuilt.

EnScript® modules, such as Initialized Case and Event Log analysis find information despite efforts to hide, cloak or delete. Easily manage large volumes of computer evidence, viewing all relevant files, including “deleted” files, file slack and unallocated space. Transfer evidence files directly to law enforcement or legal representatives as necessary. Review options allow non-investigators, such as attorneys, to review evidence with ease and reporting options enable quick report preparation.

6. WINDEX FORENSIC SOFTWARE

The following are features of this software

6.1 Disk cloning, disk imaging:

it to produce exact duplicates of disks/drives, e.g. to save the time for a full installation of the operating system and other software for several computers/disks of the same type, or to be able to restore a running installation in case of data loss/screwed up Windows (restoration of a backup). Also for computer forensics specialists, since they need to work on a copy when searching for evidence on the object disk. You can clone directly, or from an image file.

6.2 Analyzing files:

it determines the type of data recovered as lost cluster chains by ScanDisk or chkdsk.
6.3 Data recovery:
for erroneously deleted files or generally after an experienced loss of data. Can be
done manually or automatically. There is an automatic recovery mode for FAT12,
FAT16, FAT32, and NTFS drives called “File Recovery by Name” that simply requires
you to specify one or more file masks (like *.gif, John*.doc, etc.). WinHex will do the
rest. Via the Access button menu, a recovery mechanism is available for FAT drives
which re-creates entire nested directory structures (details here). Another mechanism
(“File Recovery by Type”, formerly “file retrieval”) can be used on any file system and
recovers all files of a certain type at a time. Supported file types: jpg, png, gif, tif, bmp,
dwg, psd, rtf, xml, html, eml, dbx, xls/doc, mdb, wpd, eps/ps, pdf, qdf, pw1, zip, rar, wav,
avi, ram, rm, mpg, mpg, mov, asf, mid. In particular owners of digital cameras quite often
encounter problems with their media. WinHex is likely to help with this automated
function that makes good use of the existence of file headers (characteristic signatures at
the beginning of a file). Tools | Disk Tools | File Retrieval

6.4 Computer examination/forensics:
WinHex is an invaluable tool in the hands of computer investigative specialists in
private enterprise and law enforcement.

6.5 Checksum/digest calculation:
to make sure a file is not corrupt and was not manipulated, or to identify common
known files. It is tools to calculate Hash and generating pseudo-random data.

7. FORENSIC TOOLKIT® (FTK®)

Forensic Toolkit® (FTK®) is recognized around the world as the standard in
cutting-edge computer forensic software. This court–validated digital investigations platform delivers
cutting-edge computer forensic analysis, decryption and password cracking software all
within an intuitive and customizable interface. FTK 3 is built for speed, analytics and
enterprise-class scalability. Known for its intuitive interface, email analysis, customizable
data views and stability, FTK lays the framework for seamless expansion, so your
computer forensic solution can grow with your organization’s needs.

Forensic Toolkit 3 is now the most advanced computer forensic software available,
providing functionality that normally only organizations with tens of thousands of dollars
could afford. However, we are committed to making our technology available to all
investigators and analysts, whether they are in law enforcement, education, a government
agency, a Fortune 500 corporation, or performing digital investigations as a computer
forensics service provider.

An Integrated Computer Forensics Solution, create images, analyze the registry,
conduct an investigation, decrypt files, crack passwords, identify steganography, and
build a report all with a single solution. Recover passwords from 100+ applications;
harness idle CPUs across the network to decrypt files and perform robust dictionary
attacks. KFF has hash library with 45 million hashes.
8. SMART FORENSICS

SMART is the next generation of computer forensic software – and is the clear choice of forensic investigators that want control over their forensic processes and examinations. SMART enables forensic investigators, examiners, technicians, analysts and information security professionals to quickly and efficiently conduct examinations of targeted digital data. SMART is designed to promote and facilitate efficiency throughout the forensic examination process. ASR Data designed SMART so it does not limit the options available to the end-user. Our mission is to build smart tools for smart people and then let them decide how to best employ the tools depending on the requirements of their job.

9. COMPARISON OF FORENSIC SOFTWARES

We have run WinHex, FTK, EnCase & SMART Forensic software to see their suitability, according to nature and volume and gravity of problems, whichever you choose to use very much depends on the circumstances of the case. The following are the main criteria to judge the suitability of software programs for forensic investigation.

9.1 Imaging Capability

All of the tools have imaging capabilities, but we like using SMART for this purpose. WinHex has a DOS based Replica tool which provides a true image of the target. FTK and EnCase create images which are not true copies as they embed the image files with error checking data. These three work in a DOS/Windows environment and unless careful it may be possible to miss the fact that you are not acquiring the entire contents of the drive. SMART is a LINUX tool and creates a true bit image copy of the target and can create a separate file for the error checking purposes. If you try SMART you will be pleased with the available options and the fact that you can simultaneously create multiple image files as well as clone the target.

It should be noted that these tools can all open image files created by the other tools... however, it makes sense that proprietary formats should only really be worked on using the software that created it. Using raw (true) images such as those created with SMART can be opened with any of the tools and there is no need to worry about proprietary formats.

9.2 Searching Capability

The best tool of the four (in our experience) for keyword searching is FTK. FTK can create a keyword index of the entire image at the start of the process which makes futures searches easy. It is rare that you start a case with all the correct keywords... as a case develops, you often need to repeat searches with new keywords which can waste a lot of time.

9.3 Examining Complex Structures

This is where EnCase is quite outstanding - it is capable of breaking down complex file structures for examination, such as the registry files, dbx & pst files, thumbs db etc.

9.4 Carving capability

For recovering files from unallocated space, I will commonly use EnCase, but will often try WinHex to do the same thing (basically because I really like WinHex & trust it more).
9.5 Incident Response
For incident response, all of the tools can be useful, but before I start, I would use a tool under the same OS as the system being examined. i.e., I would definitely use SMART to examine a Linux machine and one of the others for a windows machine. For live incident response, we have been testing ProDiscover with some success.

9.6 Unicode interpretation
We work with a lot of foreign language systems & the ability of the software to interpret the different unicode code pages successfully is important. The only tool which has come close so far is EnCase, but it still leaves alot to guesswork.

10. CONCLUSION
If someone is working with incident response and especially if you deal with alot of Linux systems we would strongly recommend SMART software. Besides this Winhex (X-Ways forensics), is also doing well. When you have a great deal of control over what you are doing - for automated forensic work then use EnCase or FTK software Lastly, do not overlook Pro Discover - very useful for the type of work you seem to be involved in. In addition, we used tools such as Gargoyle (from Wetstone) which is useful for quickly identifying hacking tools or other malware on a system under investigation. Paraben e-mail examiner (both local and server versions) is also very useful where e-mail is part of your investigations. But in a nutshell, all tools have their uses but just one will not meet every requirement. In case of financial constraint we would probably recommend SMART as well as Pro Discover software if you've got more money than gets one of EnCase or FTK software.

11. REFERENCES
## FORENSIC SOFTWARE TOOLS FOR WINDOWS

<table>
<thead>
<tr>
<th>Software</th>
<th>Description</th>
<th>License</th>
<th>Homepage</th>
</tr>
</thead>
<tbody>
<tr>
<td>dd for Windows</td>
<td>dd but for Windows.</td>
<td>GPL</td>
<td>Download Page</td>
</tr>
<tr>
<td>Encase 4</td>
<td>EnCase 4 is a complete forensic toolkit that covers much of the work that the I&amp;TM Forensic Analysts carry out. EnCase is the Primary I&amp;TM forensic tool</td>
<td>Commercial</td>
<td>Download Page</td>
</tr>
<tr>
<td>FTK</td>
<td>The AccessData Forensic Toolkit (FTK) is another complete forensic toolkit. FTK is recognized as one of the leading forensic tool to perform e-mail analysis.</td>
<td>Commercial</td>
<td>Download Page</td>
</tr>
<tr>
<td>MD5</td>
<td>Toast MD5 Hashing algorithm</td>
<td>GPL</td>
<td>Download Page</td>
</tr>
<tr>
<td>ISOBuster</td>
<td>IsoBuster is a CD/DVD and (Disk) Image File data recovery tool, that can read and extract files, tracks and sessions from CD-i, VCD, SVCD, CD-ROM, CD-ROM XA, DVD, DVCD and others. It also supports the following image file formats: *.DAO (Duplicator), *.TAO (Duplicator), *.ISO (Nero, BlindRead, Creator), *.BIN (CDRWin), *.IMG (CloneCD), *.CIF (Creator), *.FCD (Uncompressed), *.NRG (Nero), *.GCD (Prassi), *.P01 (Toast), *.C2D (WinOnCD), *.CUE (CDRWin), *.CIF (DiscJuggler), *.CD (CD-i OptImage) and <em>.GI (Prassi PrimoDVD). The program uses several retry-mechanisms to aid you in getting the data, even if Windows is not able to do so. Additional features include Mpg (</em>.dat) Extraction, support for file system properties, CDTex support and much more. The vast majority of the features available are free; however some advanced features like UDF support are only available in a registered version. You can choose at install time, which version to use.</td>
<td>Shareware</td>
<td>Download Page</td>
</tr>
<tr>
<td>MD5 &amp; Hashing Utilities</td>
<td>MD5 hashing algorithm</td>
<td>Shareware</td>
<td>Download Page</td>
</tr>
<tr>
<td>Software</td>
<td>Description</td>
<td>License</td>
<td>Homepage</td>
</tr>
<tr>
<td>-------------------------</td>
<td>-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
<td>-----------</td>
<td>------------------------</td>
</tr>
<tr>
<td><strong>P2 Power Pack</strong></td>
<td>This product currently contains the following items from Paraben Forensics: • Case Agent Companion v1.0; • Decryption Collection Enterprise v2.5; • E-mail Examiner v4.01; • Forensic Replicator v3.1; • Forensic Sorter v1.0; • Network E-mail Examiner v1.9; • PDA Seizure v3.0.1.35; • Text Searcher v1.0; • Chat Examiner v1.0.</td>
<td>Commercial</td>
<td>Download Page</td>
</tr>
<tr>
<td><strong>Paraben Case Agent Companion</strong></td>
<td>Paraben’s Case Agent Companion is designed to optimize both the time of the examiner and the agent working the case. Built in viewers for over 225 file formats and compatible with Paraben’s P2.</td>
<td>Commercial</td>
<td>Download Page</td>
</tr>
<tr>
<td><strong>Paraben Email Examiner</strong></td>
<td>Paraben’s E-mail Examiner is one of the most comprehensive e-mail examination tools available. E-mail Examiner claims to recover more active and deleted mail messages than the leading competitor.</td>
<td>Commercial</td>
<td>Download Page</td>
</tr>
<tr>
<td><strong>Paraben Network Email Examiner</strong></td>
<td>Network E-mail Examiner allows the user to thoroughly examine a variety of network e-mail archives. Network E-mail Examiner is designed to work hand-in-hand with E-mail Examiner and all output is compatible and can easily be loaded for more complex tasks.</td>
<td>Commercial</td>
<td>Download Page</td>
</tr>
<tr>
<td><strong>Paraben Forensic Replicator</strong></td>
<td>Replicate exact copies of drives and media. Paraben’s Forensic Replicator can acquire a wide range of electronic media from a floppy to a hard disk. Forensic Replicator images can be compressed and segmented and easily read into the most popular forensic analysis programs.</td>
<td>Commercial</td>
<td>Download Page</td>
</tr>
<tr>
<td><strong>Paraben Forensic Sorter</strong></td>
<td>Manage your data effectively and efficiently. Forensic Sorter classifies data into over 14 different categories, recovers deleted files, and filters out common hashes (FOCH), making examinations easier to manage, faster to process, and easier to find what you’re looking for.</td>
<td>Commercial</td>
<td>Download Page</td>
</tr>
<tr>
<td><strong>Paraben NetAnalysis</strong></td>
<td>Interrogates internet cache and history with powerful searching, filtering and evidence identification.</td>
<td>Commercial</td>
<td>Download Page</td>
</tr>
<tr>
<td>Software</td>
<td>Description</td>
<td>License</td>
<td>Homepage</td>
</tr>
<tr>
<td>---------------------------</td>
<td>---------------------------------------------------------------------------------------------------------------------------------------------------</td>
<td>-----------</td>
<td>---------------------</td>
</tr>
<tr>
<td><strong>Paraben Text Searcher</strong></td>
<td>Paraben's Text Searcher is a fast, comprehensive, and feature-rich text searching tool.</td>
<td>Commercial</td>
<td>Download Page</td>
</tr>
<tr>
<td><strong>SafeBack</strong></td>
<td>SafeBack is used to create mirror-image (bit-stream) backup files of hard disks or to make a mirror-image copy of an entire hard disk drive or partition.</td>
<td>Commercial</td>
<td>Download Page</td>
</tr>
<tr>
<td><strong>SHA verify</strong></td>
<td>SHA verify is a hashing program which will calculate the MD5 (128 bit), SHA1 (160 bit), SHA2 (256 bit), SHA2 (384 bit), and SHA2 (512 bit) hashes of files. A 2004 enhancement is that if you have a number of dd (flat) images, it can perform the hashes on the entire set of files and provides a single hash as if it was a single file. This is useful for confirming the hash of a physical drive against the set of dd files.</td>
<td>Freeware</td>
<td>Download Page</td>
</tr>
<tr>
<td><strong>UTK</strong></td>
<td>The Ultimate Toolkit is the complete AccessData Software Kit. This contains the FTK, DNA and PRTK.</td>
<td>Commercial</td>
<td>Download Page</td>
</tr>
<tr>
<td><strong>WinHex</strong></td>
<td>WinHex is a universal hexadecimal editor. WinHex is often used in forensic examinations</td>
<td>Freeware</td>
<td>Download Page</td>
</tr>
</tbody>
</table>
DATA QUALITY MANAGEMENT USING DATA MINING TECHNIQUES

Irshad Ullah
Department Of Computer Science, Lahore University of Management Sciences
Lahore, Pakistan. Email: irshadullah79@gmail.com

ABSTRACT

Data quality is the reliability and effectiveness of data, particularly in huge data storage. Data quality assurance (DQA) is the process of verifying the reliability and usefulness of data. Maintaining data quality requires going through the data periodically and scrubbing it. Typically this involves updating it, standardizing it, and de-duplicating records to create a single view of the data, even if it is stored in multiple disparate systems.

In this research, Data Mining approaches will be used for the data quality management process to prove experimentally and practically that how reliable, efficient and fast are these for the analysis of data? A mathematical threshold is set to analyze the data. The obtained results will be tested by applying the approach to the databases and data warehouses of different sizes with different threshold values. The results produced will be of different magnitude from short to the largest sets of data items. By this, we may take the results produced for different purpose.

KEYWORDS:
Results; Approach; Database; Process; Management.

I. INTRODUCTION

A. Data Quality Management

Simply put, data quality management entails the establishment and deployment of roles, responsibilities, policies, and procedures concerning the acquisition, maintenance, dissemination, and disposition of data. A partnership between the business and technology groups is essential for any data quality management effort to succeed. The business areas are responsible for establishing the business rules that govern the data and are ultimately responsible for verifying the data quality. The Information Technology (IT) group is responsible for establishing and managing the overall environment – architecture, technical facilities, systems, and databases – that acquire, maintain, disseminate, and dispose of the electronic data assets of the organization.

Organizations of all kinds make decisions and service customers based on the data they have at their disposal. A data warehouse is often used to examine business trends to establish a strategy for the future; within the scope of a customer relationship management (CRM) program, data about the customer is used to make appropriate decisions concerning that customer; and data in the financial systems is used to understand the profitability of past actions. The viability of the business decisions is
contingent on good data, and good data is contingent on an effective approach to data quality management 4].

**B. Data Mining**

A branch of computer science, is the process of extracting patterns from large data sets by combining methods from statistics and artificial intelligence with database management. Data mining is seen as an increasingly important tool by modern business to transform data into business intelligence giving an informational advantage. It is currently used in a wide range of profiling practices, such as marketing, surveillance, fraud detection, and scientific discovery.

The related terms data dredging, data fishing and data snooping refer to the use of data mining techniques to sample portions of the larger population data set that are (or may be) too small for reliable statistical inferences to be made about the validity of any patterns discovered. These techniques can, however, be used in the creation of new hypotheses to test against the larger data populations. [1]

### 2. LITERATURE REVIEW

**A. The AIS algorithm**

AIS Algorithm, first introduced in 1993 [5] is used to identify relationships among a set of items in a database. These relationships are not based on inherent properties of the data themselves (as with functional dependencies), but rather based on co-occurrence of the data items.

The AIS algorithm was the first published algorithm developed to generate all large itemsets in a transaction database. It focused on the enhancement of databases with necessary functionality to process decision support queries. This algorithm was targeted to discover qualitative rules. This technique is limited to only one item in the consequent. The AIS algorithm makes multiple passes over the entire database. During each pass, it scans all transactions. In the first pass, it counts the support of individual items and determines which of them are large or frequent in the database. Large itemsets of each pass are extended to generate candidate itemsets. After scanning a transaction, the common itemsets between large itemsets of the previous pass and items of this transaction are determined. These common itemsets are extended with other items in the transaction to generate new candidate itemsets. A large itemset is extended with only those items in the transaction that are large and occur in the lexicographic ordering of items later than any of the items in I. To perform this task efficiently, it uses an estimation tool and pruning technique. The estimation and pruning techniques determine candidate sets by omitting unnecessary itemsets from the candidate sets. Then, the support of each candidate set is computed. Candidate sets having supports greater than or equal to min support are chosen as large itemsets. These large itemsets are extended to generate candidate sets for the next pass. This process terminates when no more large itemsets are found. The main problem of the AIS algorithm is that it generates too many candidates that later turn out to be small [6].
B. Apriori Algorithm

Apriori scans the entire database in each pass to count support. Scanning of the entire database may not be needed in all passes. Based on this conjecture, [6] proposed another algorithm called Apriori-TID. Similar to Apriori, Apriori-TID uses the Apriori’s candidate generating function to determine candidate itemsets before the beginning of a pass. The main difference from Apriori is that it does not use the database for counting support after the first pass. Rather, it uses an encoding of the candidate itemsets used in the previous pass denoted by k C. The advantage of using this encoding function is that in later passes the size of the encoding function becomes smaller than the database, thus saving much reading effort.

The Off-line Candidate Determination (OCD) technique is proposed in [2] based on the idea that small samples are usually quite good for finding large itemsets. The OCD technique uses the results of the combinatorial analysis of the information obtained from previous passes to eliminate unnecessary candidate sets. It is obvious that if the database is very large, it is important to make as few passes over the data as possible. A good approximation of the large itemsets can be obtained by analyzing only small samples of a large database [7] [2]. Theoretical analysis performed by Mannila1994 shows that small samples are quite good for finding large itemsets. It is also mentioned in [2] that even for fairly low values of support threshold, a sample consisting of 3000 rows gives an extremely good approximation in finding large itemsets.

3. ASSOCIATION RULE PROBLEM

The following is the formal problem definition. Let I = {i1, i2,......, in} be a set of literals called items. Let D be a set of transaction, where each transaction is a set of item such that associated with each transaction is a unique identifier called TID. We say that a transaction contain T contain X, a set of some item in I, if an association rule is an implication of the form   where and X ∩ Y= . The rule hold in the set D with confidence c if c% of transaction in D that contain X also contain Y. The rule has confidence S in the transaction set D if s% of transaction in D contain XUY [5].

Association rules can be classified based on the type of vales, dimensions of data, and levels of abstractions involved in the rule. If a rule concerns associations between the presence or absence of items, it is called Boolean association rule. And the dataset consisting of attributes which can assume only binary (0-absent, 1-present) values is called Boolean database.

III Binary Variables

A binary variable has only two states: 0 or 1, where 0 means that the variable is absent, and 1 means that it is present. If all binary variables are thought of as having the same weight, we have the 2-by-2 contingency table of table 4.4, where q is the number of variables that equal 1 for both items i and j, r is the number of variables that equal 1 for item i but that are 0 for item j, s is the number of variables that equal 0 for item i but equal 1 for item j, and t is the number of variables that equal 0 for both item i and j. The total number of variables is p, where p = q + r + s + t.
Table 1: Contingency Table

<table>
<thead>
<tr>
<th>Item i</th>
<th>Item j</th>
<th>1</th>
<th>0</th>
<th>Sum</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Q</td>
<td>r</td>
<td></td>
<td>q + r</td>
</tr>
<tr>
<td>0</td>
<td>S</td>
<td>t</td>
<td></td>
<td>s + t</td>
</tr>
<tr>
<td>Sum</td>
<td>q + s</td>
<td>r + t</td>
<td>p</td>
<td></td>
</tr>
</tbody>
</table>

For noninvariant similarities, the most well-known coefficient is the Jacquard dissimilarity coefficient, where the number of negative matches \( t \) is considered unimportant and thus is ignored in the computation:

\[
d(i, j) = \frac{r + s}{q + r + s}
\]

The measurement value 1 suggests that the objects \( i \) and \( j \) are dissimilar and the measurement value 0 suggests that the objects are similar.

4. RESULTS

Two algorithms SI and Apriori were used for the implementation purpose. Different experiments were performed to check the results and accuracy of the algorithms. The algorithm requires the data in the database to be in binary format. The dataset is downloaded from the net. The data was stored in a text file on the 3.

After loading the data into the algorithms are implemented on the database having forty thousand records. The results produced by the algorithms were same. First the experiment is performed with (SI) algorithm. For the experiment input threshold was 80% (dissimilarity). The results generated by the algorithm on the given support are given below.

<table>
<thead>
<tr>
<th>Dissimilarity</th>
<th>Threshold</th>
<th>Status</th>
</tr>
</thead>
<tbody>
<tr>
<td>.71</td>
<td>.8</td>
<td>Item1 and item2 are frequent</td>
</tr>
<tr>
<td>.38</td>
<td>.8</td>
<td>Item1 and item3 are frequent</td>
</tr>
<tr>
<td>.33</td>
<td>.8</td>
<td>Item1 and item4 are frequent</td>
</tr>
<tr>
<td>.73</td>
<td>.8</td>
<td>Item1 and item5 are frequent</td>
</tr>
<tr>
<td>.64</td>
<td>.8</td>
<td>Item2 and item3 are frequent</td>
</tr>
<tr>
<td>.71</td>
<td>.8</td>
<td>Item2 and item4 are frequent</td>
</tr>
<tr>
<td>.79</td>
<td>.8</td>
<td>Item2 and item5 are frequent</td>
</tr>
<tr>
<td>.6</td>
<td>.8</td>
<td>Item3 and item4 are frequent</td>
</tr>
<tr>
<td>.75</td>
<td>.8</td>
<td>Item1, item2 and item3 are frequent</td>
</tr>
<tr>
<td>.53</td>
<td>.8</td>
<td>Item1, item3 and Item4 are frequent</td>
</tr>
</tbody>
</table>

Figure 1: SI algorithm results
<table>
<thead>
<tr>
<th>Support</th>
<th>Support Count</th>
<th>Status</th>
</tr>
</thead>
<tbody>
<tr>
<td>20</td>
<td>9676</td>
<td>Item1 is Frequent</td>
</tr>
<tr>
<td>20</td>
<td>7736</td>
<td>Item2 is Frequent</td>
</tr>
<tr>
<td>20</td>
<td>10 664</td>
<td>Item3 is Frequent</td>
</tr>
<tr>
<td>20</td>
<td>9696</td>
<td>Item4 is Frequent</td>
</tr>
<tr>
<td>20</td>
<td>8688</td>
<td>Item5 is Frequent</td>
</tr>
<tr>
<td>20</td>
<td>3876</td>
<td>Item1, Item2 are Frequent</td>
</tr>
<tr>
<td>20</td>
<td>7756</td>
<td>Item1, Item3 are Frequent</td>
</tr>
<tr>
<td>20</td>
<td>7756</td>
<td>Item1, Item4 are Frequent</td>
</tr>
<tr>
<td>20</td>
<td>4846</td>
<td>Item2, Item3 are Frequent</td>
</tr>
<tr>
<td>20</td>
<td>3876</td>
<td>Item2, Item4 are Frequent</td>
</tr>
<tr>
<td>20</td>
<td>6786</td>
<td>Item3, Item4 are Frequent</td>
</tr>
<tr>
<td>20</td>
<td>4848</td>
<td>Item4, Item5 are Frequent</td>
</tr>
<tr>
<td>20</td>
<td>3876</td>
<td>Item1, Item2, Item3 are Frequent</td>
</tr>
<tr>
<td>20</td>
<td>6786</td>
<td>Item1, Item3, Item4 are Frequent</td>
</tr>
</tbody>
</table>

Figure 2: SI algorithm results.

By changing size of the database and also the input threshold the results generated by the algorithms were same.

After this it’s now clear that this is a reliable technique for the enhancement of the quality of data.

5. CONCLUSION AND FUTURE WORK

The experiments showed the performance and effectiveness of the technique. In future we intend to compare this algorithm with the established algorithms of association rule mining for efficiency purpose.

6. REFERENCES


TECHNIQUES USED FOR HACKING THE INFORMATION

Nadia Qasim\textsuperscript{1} and Muhammad Qasim Rind\textsuperscript{2}

\textsuperscript{1} Kings College London, Email: nadia.nadya@gmail.com
\textsuperscript{2} Preston University, Islamabad, Pakistan. Email: dr_rind@yahoo.com

ABSTRACT
The paper shows full history of hacking demonstrates that hacking is increasing tremendously, with increasing use of internet. The hacker ethic imply that it is the ethical duty of hackers to protect information, respect others privacy, use the resources that are being wasted, respect peoples' right to correspond with each other, Share the information, etc. Always test security and system integrity of computer systems. Hacker culture(s) today come out of hobby hacking, academic hacking and network hacking. It is more or less based on an ethical code, interpreted and shared in different ways. By understanding their development and motivations, it is hoped that elements leading to the development and building of motivations of hackers would be checked and controlled.

KEYWORDS
Hackers; Hacker Ethics; Hackers Culture; Academic Hacking and network hacking.

1. INTRODUCTION
To most of the computing world, especially for programmers, there is a certain awe and fascination for hackers. Hackers are regarded with a certain degree of respect, not because of their acts of violations but because of their intellectual prowess. By understanding their development and motivations, it is hoped that elements leading to the development and building of motivations of hackers would be checked and controlled.

The full history of hacking that includes 1960’s, 1970’s, 1980’s, 1990’s, 2000’s hacker first have to go through lot of different stages like mundane person, lamer, wannabe, larva, etc to became a real hacker. Hackers’ ethics have ethical principles of the original.

Hacker ethic that it is information wants to be free, mistrust authority, hands on imperative; computers can change your life for the better, etc. Hackers’ ethics have new hacker ethic that includes do no harm, protect privacy waste not, want not, self defense, hacking helps security, etc. There are different types of hacker like vandals, gamers, spies, etc. Study different Hacker culture(s), methods used to hack into computers.

2. HACKERS AND HACKING
Concept of hacking first came into picture in the 1970s and was introduced by Massachusetts Institute of Technology (MIT). Hacking can be defined as when any person attempts to breach the security of remote computers. Originally, this term comes from a community of expert programmers and networking wizards that traces its history
Techniques used for hacking the information

back through decades to the first time-sharing minicomputers and the earliest ARPAnet experiments. The members of this culture gave birth to the term “hacker”. Many experts said that Hackers have contributed to the development of the Internet and operating systems such as UNIX or Usenet.

A person having tremendous knowledge about computers and who is capable of making the computer do anything is a hacker. They are very tolerant and inquisitive people. The prime reason for the hackers to breach the security of the remote computer and get into it is not to damage the system or to sell what they obtain but to gain knowledge and to gratify their inquisitiveness (hamilton.edu, 2003).

The hacker gets into the remote system by creating a hack. A hack can be defined as either a prank or a program which can break into other persons system. A good hack should be a perfectly laid out program, which is successful, and accomplishes the goal of hacking (hamilton.edu, 2003).

3. HISTORY OF HACKING

Hacking has been around pretty much since the development of the first electronic computers. Full history of hacking demonstrates that hacking is increasing tremendously, with increasing use of internet as shown following (security metrics, 2002):

Here are some of the key events in the last four decades of hacking (evidian, 2004).

3.1 (1960s): The Dawn of Hacking

The first computer hackers emerge at MIT. They borrow their name from a term to describe members of a model train group at the school who "hack" the electric trains, tracks, and switches to make them perform faster and in different ways. A few of the members transfer their curiosity and rigging skills to the new mainframe computing systems being studied and developed on campus.

3.2 (1970s): Phone Phreaks and Cap'n Crunch

Phone hackers (phreaks) break into regional and international phone networks to make free calls. One phreak, John Draper (aka Cap'n Crunch), learns that a toy whistle given away inside Cap'n Crunch cereal generates a 2600-hertz signal, the same high-pitched tone that accesses AT&T's long-distance switching system. Draper builds a "blue box" that, when used in conjunction with the whistle and sounded into a phone receiver, allows phreaks to make free calls. Shortly thereafter, Esquire magazine publishes "Secrets of the Little Blue Box" with instructions for making a blue box, and wire fraud
in the United States escalates. Among the perpetrators: college kids Steve Wozniak and Steve Jobs, future founders of Apple Computer, who launched a home industry making and selling blue boxes.

3.3 (1980): Hacker Message Boards and Groups

Phone phreaks begin to move into the realm of computer hacking, and the first electronic bulletin board systems (BBSs) spring up. The precursor to Usenet newsgroups and e-mail, the boards—with names such as Sherwood Forest and Catch-22—become the venue of choice for phreaks and hackers to gossip, trade tips, and share stolen computer passwords and credit card numbers.

Hacking groups begin to form. Among the first are Legion of Doom in the United States, and Chaos Computer Club in Germany.

3.4 (1983): Kids' Games

The movie War Games introduces the public to hacking, and the legend of hackers as cyber heroes (and anti-heroes) is born. The film's main character, played by Matthew Broderick, attempts to crack into a video game manufacturer's computer to play a game, but instead breaks into the military's nuclear combat simulator computer. The computer (codenamed WOPR, a pun on the military's real system called BURGR) misinterprets the hacker's request to play Global Thermonuclear War as an enemy missile launch. The break-in throws the military into high alert, or Def Con 1 (Defense Condition 1). The same year, authorities arrest six teenagers known as the 414 gang. During a nine-day spree, the gang breaks into some 60 computers, among them computers at the Los Alamos National Laboratory, which helps develop nuclear weapons.

3.5 (1984): Hacker 'Zines

The hacker magazine 2600 begins regular publication, followed a year later by the online 'zine Phrack. The editor of 2600, "Emmanuel Goldstein" (whose real name is Eric Corley), takes his handle from the main character in George Orwell's 1984. Both publications provide tips for would-be hackers and phone phreaks, as well as commentary on the hacker issues of the day. Today, copies of 2600 are sold at most large retail bookstores.

3.6 (1986): Use a Computer, Go to Jail

In the wake of an increasing number of break-ins to government and corporate computers, Congress passes the Computer Fraud and Abuse Act, which makes it a crime to break into computer systems. The law, however, does not cover juveniles.

3.6 (1988): The Morris Worm

Robert T. Morris, Jr., a graduate student at Cornell University and son of a chief scientist at a division of the National Security Agency, launches a self-replicating Worm on the government's ARPAnet (precursor to the Internet) to test its effect on UNIX systems. The worm gets out of hand and spreads to some 6000 networked computers, clogging government and university systems. Morris is dismissed from Cornell, sentenced to three years' probation, and fined $10,000.
3.7 (1989): The Germans and the KGB

In the first cyber espionage case to make international headlines, hackers in West Germany (loosely affiliated with the Chaos Computer Club) are arrested for breaking into U.S. government and corporate computers and selling operating-system source code to the Soviet KGB. Three of them are turned in by two fellow hacker spies, and a fourth suspected hacker commits suicide when his possible role in the plan is publicized. Because the information stolen is not classified, the hackers are fined and sentenced to probation. In a separate incident, a hacker is arrested who calls himself The Mentor. He publishes a now-famous treatise that comes to be known as the Hacker's Manifesto. The piece, a defense of hacker antics, begins, "My crime is that of curiosity ... I am a hacker, and this is my manifesto. You may stop this individual, but you can't stop us all."

3.8 (1990): Operation Sun devil

After a prolonged sting investigation, Secret Service agents swoop down on hackers in 14 U.S. cities, conducting early-morning raids and arrests. The arrests involve organizers and prominent members of BBSs and are aimed at cracking down on credit-card theft and telephone and wire fraud. The result is a breakdown in the hacking community, with members informing on each other in exchange for immunity.

3.9 (1993): Why Buy a Car When You Can Hack One?

During radio station call-in contests, hacker-fugitive Kevin Poulsen and two friends rig the stations' phone systems to let only their calls through, and "win" two Porsches, vacation trips, and $20,000. Poulsen, already wanted for breaking into phone-company systems, serves five years in prison for computer and wire fraud. (Since his release in 1996, he has worked as a freelance journalist covering computer crime). The first Def Con hacking conference takes place in Las Vegas. The conference is meant to be a one-time party to say good-bye to BBSs (now replaced by the Web), but the gathering is so popular it becomes an annual event.

3.10 (1994): Hacking Tools

The Internet begins to take off as a new browser, Netscape Navigator, makes information on the Web more accessible. Hackers take to the new venue quickly, moving all their how-to information and hacking programs from the old BBSs to new hacker Web sites. As information and easy-to-use tools become available to anyone with Net access, the face of hacking begins to change.

3.11 (1995): The Mitnick Takedown

Serial cyber trespasser Kevin Mitnick is captured by federal agents and charged with stealing 20,000 credit card numbers. He's kept in prison for four years without a trial and becomes a cause célèbre in the hacking underground. After pleading guilty to seven charges at his trial in March 1999, he's eventually sentenced to little more than time he had already served while he wait for a trial. Russian crackers siphon $10 million from Citibank and transfer the money to bank accounts around the world. Vladimir Levin, the 30-year-old ringleader, uses his work laptop after hours to transfer the funds to accounts in Finland and Israel. Levin stands trial in the United States and is sentenced to three years in prison. Authorities recover all but $400,000 of the stolen money.
3.12 (1997): Hacking AOL
AOHell is released, a freeware application that allows a burgeoning community of unskilled hackers--or script kiddies--to wreak havoc on America Online. For days, hundreds of thousands of AOL users find their mailboxes flooded with multi-megabyte mail bombs and their chat rooms disrupted with spam messages.

The hacking group Cult of the Dead Cow releases its Trojan horse program, Back Orifice--a powerful hacking tool--at Def Con. Once a hacker installs the Trojan horse on a machine running Windows 95 or Windows 98, the program allows unauthorized remote access of the machine. During heightened tensions in the Persian Gulf, hackers touch off a string of break-ins to unclassified Pentagon computers and steal software programs. Then-U.S. Deputy Defense Secretary John Hamre calls it "the most organized and systematic attack" on U.S military systems to date. An investigation points to two American teens. A 19-year-old Israeli hacker who calls himself The Analyzer (aka Ehud Tenebaum) is eventually identified as their ringleader and arrested. Today Tenebaum is chief technology officer of a computer consulting firm.

In the wake of Microsoft's Windows 98 release, 1999 becomes a banner year for security (and hacking). Hundreds of advisories and patches are released in response to newfound (and widely publicized) bugs in Windows and other commercial software products. A host of security software vendors release anti-hacking products for use on home computers.

In one of the biggest denial-of-service attacks to date, hackers launch attacks against eBay, Yahoo, Amazon, and others. Activists in Pakistan and the Middle East deface Web sites belonging to the Indian and Israeli governments to protest oppression in Kashmir and Palestine. Hackers break into Microsoft's corporate network and access source code for the latest versions of Windows and Office.

3.16 (2001): DNS Attack
Microsoft becomes the prominent victim of a new type of hack that attacks the domain name server. In these denial-of-service attacks, the DNS paths that take users to Microsoft's Web sites are corrupted. The hack is detected within a few hours, but prevents millions of users from reaching Microsoft Web pages for two days.

3.17 (2002): NASA Attacks
Hacker case increased, second offense hacked into NASA computers. Former employee of The Bergman Companies (TBC) gets unauthorized e-mail access for commercial gain and caused approximately $21,636 in damages and costs to TBC. Unauthorized access of credit card accounts at defraud Priceline.com and others with credit card information unlawfully obtained from a credit union employee and causing damage more than $116,000. British National hacked into U.S. Military Networks.
3.18 (2003): Redirect Web Traffic
Hacking activates include contract employee attacked NASA e-mail server. Website designer redirected Aljazeera.net web traffic. Denial-of-service attacks on many Web sites increased. FBI employee exceeded authorized access on FBI computer.

4. FAMOUS HACKERS OF WORLD

Here is the short list of the 4 most famous hackers in history (evidian, 2004):

4.1 Robert Morris
- 24 years old - Cornell University Grad who released a computer worm that crashed 6,000 computers on the internet.

4.2 Kevin Poulsen
- 23 years old - Hacked into a radio station phone system and guaranteed that he would be the 120th caller, winning him a new Porsche.

4.3 Mark Abene
(Phiber Optik) - 22 years old - Founding member of Masters of Deception one of the leading hacking groups in the late 80s.

4.4 Kevin Mitnick
- First hacker to be put on the FBI's Most wanted list.

5. HACKER STAGES

A human being who aims to become a hacker will typically pass through these stages (dvara, 2003):

5.1 Mundane person
He basically doesn't know anything about the hacking scene, even if he may have a computer and Internet access. The only thing he knows about hackers is that they break computer systems and are criminals. Some of them write for the newspapers.

5.2 Lamer
One who confuses the hacking scene with different realities such as the warez scene. He has a very poor knowledge of the whole thing, and tries to impress mundane people with big words. His greatest achievement is to put a Trojan (wrote by someone else, and of which he is totally clueless about how it works) in someone else's computer during an IRC or ICQ chat and delete their files. People that succeed in becoming hackers usually pass this stage very quickly, or might skip it at all.

5.3 Wannabe
A Wannabe hacker found out that hacking is much more than breaking into someone else's computer and it's rather a philosophy, or a way of life. He just wants to know more, and starts to read hacking tutorials, and searches the Net for serious hacking-related stuff.

5.4 Larva
Also referred as Newbie, a hacker in his larval stage learns the basic techniques of hacking, discovers his first’s exploits, and might try to break into someone else's system,
just to make sure he figured how to do that. However, at this stage he knows he shouldn't damage the system nor delete anything, if it's not strictly necessary to cover his tracks.

5.5 Hacker

It is hard to say when the final stage of hacker has been reached, since there's always something new to learn and to discover (for collecting information and exploring the boundaries is the same essence of a hacker), but it's probably more something you feel. After all, being a hacker is more a state of mind, and if you are not born hacker, you'll never be such.

5.6 Ueberhacker

This is an unusual character, probably inspired to Nietzsche's uebermann ("overman"). It appears on a document titled "A Guide to Internet Security" by Christopher Klaus, dated December 5th, 1993, where the author suggests how to fool hackers thanks to some "social engineering", gaining thus the status of "Ueberhacker".

6. HACKER ETHICS

6.1 Original Computer Hacker Ethics

The six old but original ethics of the hacker ethic are as follows (hamilton.edu, 2003):

6.1.1 "Hands on Imperative":

According to this ethic, everyone has the right to access computers and other resources. There should not be any restraints placed on the use of computers or other resources based on age, sex, or race. Everyone should have access to computers and other resources.

6.1.2 "Information Wants to Be Free":

The "Free" in this law can mean three of the following concepts: without restrictions, without control or without any cost.

6.1.3 "Mistrust Authority":

This code promotes devolution. Hackers have always shown disbelieve toward large institutions. Tools like the PC are said to move the power away from large organizations and put them in the hands of the small, powerless, and growing organizations.

6.1.4 "No Bogus Criteria":

According to this code, the type of hacking they perform and the success of it, instead of their race, age, or sex, should judge hackers.

6.1.5 "You can create truth and beauty on a computer":

According to this code, hacker has the ability to create programs or software that is good, and helpful to people, instead of creating something that harms the users.

6.1.6 "Computers can change your life for the better":

It emphasizes the fact that since hackers have the ability to write programs, they can improve the lives of the users and make it safer.

Hence, the ethical principles of the original Hacker Ethic suggest that it is the ethical duty of the hacker to remove barriers, liberate information, distribute power, recognize
people based on their ability, and create things that are good and life-enhancing through computers.

6.2 New Computer Hacker Ethics

The hacker of the late 90s follows the new hacker ethic and lives by it. The new ethic appears to have developed like the old one, informally, and it is the continuity from the original ethic code (hamilton.edu, 2003).

6.2.1. "Above all else, does no harm":

According to this code, the hack must be secured, should not damage anything, and should not harm anyone.

6.2.2. "Protect Privacy":

According to this code, people have a "right to privacy, which means control over their own information."

6.2.3. "Waste not, want not":

According to this code, it is ethically wrong for people to be kept away from using the idle resources. Hence, computer resources should not lie idle and wasted.

6.2.4. "Exceed Limitations":

Hacking is about the continual transcendence of problem limitations.

6.2.5. "The Communicational Imperative":

According to this law, people have the right to communicate and associate with others freely. The 1st amendments" right to communication supports this law, since it is important for the free flow of information.

6.2.6. "Leave No Traces":

According to this code, no trace or trail should be left of their presence. "One should keep quiet, so everyone can enjoy what they have." This is a moral belief, since the hacker follows it for not only his selfish reasons, but also to make sure that other hackers don’t get caught. The Rules that a Hacker follows are to "keep a low profile; if suspected, then keep a lower profile; if accused, deny it; if caught plea the 5th law."

6.2.7. "Share":

Information sharing is encouraged by this law since it says that information needs to be spread by sharing it with the maximum number of people. It is not right for the information to be hidden. "The Pirates' ethic is that piracy increases interest in software, since people are given a chance to try it out and experiment with it before paying for it. So sharing software with your friends is a good thing."

6.2.8. "Self Defense":

This code is about being protected against a Cyberpunk Future Hacking and viruses.

6.2.9. "Hacking Helps Security":

Hacking can be useful and helpful to find loopholes in the security system, so it can be fixed. Hacking can be a positive force, because it shows people how to mend weak security. At times it brings forth the fact that it is impossible to have total security without making an extreme sacrifice.
6.2.10. "Trust, but Test":

It is essential to test the integrity of systems and find ways to improve them. It is necessary to know the system you are working with and not leave it to others for maintenance. "If you can exploit certain systems (such as the telephone network) in ways that their creators never intended or anticipated, that's all to the better. This could lead to creation of better systems. One of those systems that may require constant revision, testing, and adjustment, apparently, is constitutional democracy."

Thus, the new hacker ethic suggests that it is the ethical duty of new hackers to protect information, respect others' privacy, use the resources that are being wasted, exceed unnecessary restrictions, respect people's right to correspond with each other, erase all the traces, share the information, be vigilant against cyber-tyranny, always test security, and system integrity of computer systems.

7. HACKER CULTURE(s)

Hacker culture(s) today come out of hobby hacking, academic hacking, and network hacking. It is more or less based on an ethical code, interpreted and shared in different ways. There are a few dimensions that seem to span the field in a useful way (webzone.k3.mah.se, 2004). The hacker mind-set is not confined to this software-hacker culture. There are people who apply the hacker attitude to other things, like electronics or music.

According to this culture, there is another group of people who loudly call themselves hackers, but aren't. These are people who get a kick out of breaking into computers and crashing the phone system. "Original" hackers call these people "crackers" and want nothing to do with them. "Real hackers" mostly think that crackers are lazy, irresponsible, and not very bright, and object that being able to break security doesn't make you a hacker any more. The basic difference is: hackers learn to build things; crackers break them (evidian, 2004). As research shows the statistics (Zone-H.org, 2003):

<table>
<thead>
<tr>
<th>Reason</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Opportunistic – Just For Fun</td>
<td>31.3%</td>
</tr>
<tr>
<td>No Reason Specified</td>
<td>25.6%</td>
</tr>
<tr>
<td>Notoriety</td>
<td>13.8%</td>
</tr>
<tr>
<td>As A Challenge</td>
<td>10.2%</td>
</tr>
<tr>
<td>Patriotism</td>
<td>9.1%</td>
</tr>
<tr>
<td>Political Reasons</td>
<td>8.2%</td>
</tr>
<tr>
<td>Revenge Against That Website</td>
<td>1.8%</td>
</tr>
</tbody>
</table>

Source: Zone-H.org defacement Archives 1 Feb 2003

This statistic shows that the people do hacking just for fun have the highest rate of 31.3%, most of people do hacking with no specific reason have second highest rate of 25.6%. Notoriety, as a challenge, patriotism, political reasons, revenge against that website have rate of 13.8%, 10.2%, 9.1%, 8.2%, 1.8% respectively.

7.1 Hacking --- Cracking.

Real hackers are careful to point out that malicious hacking activities should properly speaking be called cracking (webzone.k3.mah.se, 2004). However, the question is where
to draw the line. The police, the corporate world, the judicial system, etc take a fairly restrictive position. Much of what hackers would call exploration for the sake of learning is regularly prosecuted.

Hacking and cracking in the late 90s has taken a few more visible forms. Defacing web pages is very popular, given the enormous visibility of the results. This basically means cracking a computer that runs a web server and places your own pages there instead of the original information. Due to the public nature of web and mail servers, they can be cracked also without access to the computer on which they run. Denial-of-service attacks on public servers, which entail sending millions of requests to the servers simultaneously from many sources, are quite frequent. Mail bombing can be seen as a variation on the same theme. Creating and disseminating a virus is another form of hacking/cracking that has taken off with the increasing penetration of Internet usage. Email is now by far the most common carriers of virus and Trojan horses.

7.2 Purpose.

Academic hacker culture views intrusion as a means for learning more about computers and networks. Another common hacker argument for exposing security flaws by intrusion is to help build safer systems in the future.

Contrary to the traditional hacker norm of keeping a low profile, many of the web defacement attacks are of the graffiti kind. Hacking/cracking has often been used as a means for personal revenge. It is not unknown for police officers investigating computer crime to receive personal credit card bills and phone bills in huge amounts. The hacker has gained access to, e.g., the phone company and manipulated the records (webzone.k3.mah.se, 2004).

8. TYPES OF HACKERS

Considering all the different types of attacks and intrusions, it is obvious that there are several types of hackers. In this short classification, we have tried to identify 4 profiles based on the nature of their actions (evidian, 2004).

a. **Vandals**: they break for fun or they destroy to shock. Vandals disfigure websites, use Logic Bomb, break up data or paralyze systems by over-flooding. Their motives can be political and/or economic or nihilist (e.g. displeased employees).

b. **Gamers**: they like challenges and technical exploits. Gamers focus on penetrating so-called "inviolable" systems and let inoffensive trails of their actions (e.g. messages, tag on websites). They correspond to the original 'hacker' meaning, in the high tech. culture

c. **Thieves**: they penetrate Websites to misappropriate funds. As regular thieves, this category is driven by profit.

d. **Spies**: they penetrate websites for industrial or political espionage. Spies try to be the most unobtrusive as possible. Spies can occasionally act like vandals in order to paralyze one company organization or economic/military forces of an opponent country.

Interestingly, whilst mostly 'kid hackers' are popularized by media, they do not represent - by far - the majority. In fact, facts show that external hackers are just the top
of the iceberg. Around 50% to 70% of break-ins and attacks do come from within the organizations, from disgruntled or malevolent employees (evidian, 2004).

9. HACKING TECHNIQUES USED BY DIFFERENT HACKERS

The methods used to hack into computers can vary widely. New hacker techniques are developed and new security vulnerabilities in networks are found every day. Hackers are getting more and more advanced and thus, harder to prevent and detect. On average, it takes less than 10 minutes to hack into even the most secure company’s servers. They can gain access quickly because most hackers will do some research about the system they are trying to gain access to first (Steve, 1998). They will try and find out anything useful about the network they are trying to break into such as the type of firewall, networking software, operating systems in use, as well as holists, usernames, network connections, and sibling domains. Often there are peer network connections that can be used as back doors into a network. In a Wired interview with Peter Shipley, the founder of Network Security Associates, he gave an example of such an incident. Shipley used the example of hackers gaining access into NASA computers, by going through Lockheed and getting in through their connections (Steve, 1998). The basic idea is for hackers to find vulnerabilities in the computer system that they can exploit. The latest vulnerabilities can be found on the COAST or CERT web servers or on any number of hacker sites. The more obvious vulnerabilities are: guessable passwords, accounts with no passwords, anonymous ftp access, word writable directories, and passwords sent in clear text across the network from one computer to another (Steve, 1998). Dan Farmer used his SATAN software (originally designed to help computer administrators detect weak parts of their system) to conduct a nonscientific study and he found security holes in about 2,200 Web sites and found about 70% to 80% with “serious security flaws (csci.ca/breach.htm, 1999).

Hacking is increased enormously because of advancement in hacking technique as shown below (Zone-H.org, 2003):
Techniques used for hacking the information

Some of the more common methods hackers use to gain access to computer systems are listed below:

9.1 Password Cracking and Theft: A hacker can steal or guess a password or encryption key in order to gain access to a computer system. Using this method, a hacker does not have to sit at the computer and guess the password, the computer can actually do the guessing itself. Cracking programs such as, Crack, and a dictionary is often all the tools that are needed. The cracking program takes each word in the dictionary, uses it as a key to encrypt the known block of data, and then compares the result with an entry in the password file. The program can also take dictionary words and try them backwards or in another common pattern. Passwords and encryption keys could also be breakable by brute force, that is, by trying all possible characters or bit combinations until one is found that works.

9.2 Packet Sniffing:
Packet sniffing is when a hacker surreptitiously inserts a software program at remote network switches or a host computer. The program then monitors the information inputted in the computer and sends it to the hacker. This way, the hacker can learn passwords and user identifications that they can then use to break into the system. Once the intruder in a system, he/she acts as a legitimate user. They are able to steal information from the compromised system as well as from any system connected to it. Thus, this can lead to a domino effect.

9.3 Jamming or Flooding:
Attacks using this method to disable or tie up a system’s resources. This is known as a "denial of service attack." An intruder can consume all the available memory or disk space on a machine and then flood it with so much traffic that no one else can use it. One way of doing this is to flood the victim’s mailbox with thousands of e-mails, thus jamming his mailbox and eventually shutting down access. Attackers can conceal their identity by using a forged return address or by directing the message through an anonymous remailer.

9.4 Sendmail:
Sendmail is a common type of attack in which the attacker installs malicious code in an electronic mail message that adds a password into the system's password file. This then gives the attacker total system privileges.

9.5 IP Spoofing:
IP spoofing means that a hacker poses as a legitimate host using a fabricated IP address, that is, it lets one computer impersonate another. This tricks the firewall into letting the intruder through the network. The intruder can then masquerade as that user throughout the system. Through IP spoofing, the intruder can use the system as a springboard to log into another system and yet another system, and so on. This process is called "looping". Inherently, it conceals the identity and location of the intruder. This can make computer intrusions hard, if not impossible to trace.
9.6 Social Engineering:

Social Engineering is a relatively non-technical technique. It is the process of learning about how computers are used by individuals and organizations. It can merely consist of a hacker calling up a network engineer, or someone else with access to information, and simply asking what type of software and configurations or port assignments are being used in their network. Social engineering is often cited as the easiest way into a system.

10. HACKERS ON THE INTERNET

Hackers cannot access anything on a computer, unless it has a program set to connect and respond to other systems. One example of something a hacker can do is access the password file on a server, and then use the accounts that they have acquired the passwords for. This can make it easy for hackers to put blame on other people, when they use accounts owned by someone else to do something illegal. It is also possible for hackers to obtain credit card numbers and use them. Hackers can also access almost any other information on an Internet server, if they are skilled enough. Hackers can easily send fake E-mail, and create fake posts to UseNet newsgroups.

This way, no one will know who sent the mail, or created the post. It is possible to create virii and distribute them, causing the destruction of data. Hackers can also crash servers, causing Internet sites to be off-line until fixed (Belisarius, 1999). Hackers can bypass many security measures on servers. Most computer systems have numerous flaws, which will almost always be discovered and exploited by hackers. This is how hackers gain access to protected information (passwords, credit cards, etc.). There are legal matters involved with hackers. Illegal computer activities fall into a few broad categories. First, unauthorized entry into a computer system is a crime; common reasons are pride, curiosity, or a sense of adventure. Second, unauthorized destruction or theft of data is a crime; this includes virii, or any tampering with a system (Judson, 2002). Hackers can be difficult to trace, and therefore hard to prosecute.

Hackers often use fake identifications, either by using someone else’s account, if they have gained access to it, or masking their own identity, by either creating a program that intercepts any attempts at discovering their identity, or destroying system logs. Also, a large part of hacking is phone hacking, or "phreaking." By phreaking, hackers can often connect to the Internet on lines other than their own. Phone hackers do not need to worry about their phone line being traced, or about long distance bills (Belisarius, 1999). When hackers do commit illegal acts, they can be hard to trace. True hackers are only a small threat, whereas crackers and other groups are becoming larger threats. True hackers that follow the hacker ethic are a very small threat. Crackers who do not follow the hacker ethic are becoming a larger threat. Another group of people that are commonly associated with hackers are warez dealers. Warez dealers are people who illegally distribute commercial software to other people. They are being seen more often, and are becoming a threat to software companies, because they freely distribute software that the companies would normally be selling (Goggans, 2001).

Company insiders are currently the biggest threat to computer security. Company insiders are people who work in a company. They know the systems well, and have access to the systems that would probably not be questioned. In general, company
Techniques used for hacking the information

Insiders are a much larger threat than the hackers and crackers who are capable of breaking into an advanced system. It is said that less than 100 hackers are skilled enough to penetrate high security systems and actually worry the system administrator (Judson, 2002).

The following chart indicates the type of attacker and what kind of attack they are likely to do.

![Chart showing types of attackers and their attacks](Hale, 2002)

Most hackers are merely curious to see if they can get into the system, but plan on doing no harm once they are inside. Others however, hack to cause harm. Some vandalize web sites while others actually steal information, software or money.

11. HOW TO PROTECT AGAINST HACKING?

By knowing your enemy you can at least raise the odds of not becoming the victim of a hacker (evidian, 2004). Here are 10 golden rules in setting up bullet proof security:

10.1 Define a security policy

Each organization is unique. Their business activities, resources, and organizational structure call for appropriate security measures. The first step consists of analyzing the context in order to identify the possible choices. It is possible to weigh up solutions properly, if a preliminary risk analysis is carried out methodically. What would be the cost to me if my client file was unavailable, lost, or even leaked? Simple scenarios like this can be used to identify needs, which in turn can be used to define priorities and areas where intervention is requires.

10.2 Identify and audit internal security requirements

A precise evaluation of potential weaknesses is essential and this applies both internally and externally. 70% of break-ins are actually accomplished with the help of an insider. Whether they originate from the thrill of risk-taking or real abuse, these attacks often have devastating effects. The workplace is becoming less and less a unifies area of trust. Security policy must take this into account by providing effective protection for sensitive servers, and by partitioning trust along functional and/or hierarchical lines.

10.3 Take account of the risks inherent in opening up networks

Networks create significant security problems. Organization boundaries are shifting: new subsidiaries are added, electronic transactions with suppliers are possible, customer
access is provided, etc. The organization's expansion is dependent on the growth of its interconnections. However, this new openness must be controlled, perhaps by partitioning services and securing data flows. The new trends in architecture, like the development of mobile code (Java applets, ActiveX components…), also carry risks that have to be taken into consideration. In this framework, openness and the standardization of application exchange protocols make it possible to implement security control.

10.4 Keep up-to-date with changes in legislation

In the field of computer security it is vital to keep up to date with national and international legislation. Legislation in this field is often highly restrictive, especially concerning data encryption. It is also important to understand the implications of Data Protection lax, as this can frequently have implications for data integrity and access control.

10.5 Create dedicated security structures

Security management requires dedicated teams with precise assignment of responsibilities. It is essential to set up a centralized organization devoted to security. It is also useful to deploy local security committees to take decisions on logical security, application access, and also physical access. These committees will allow fine-tuning the organization's security management.

10.6 Choose solutions that conform to standards and allow interoperability

Openness and standards compliance are important, for the approach must be global. Companies have long perceived security as an expensive, complicated process, Instead of ensuring their continued future existence and prosperity, this attitude has inhibited the use of new working methods. However, tried and tested security solutions already exist to enable enterprises to take advantage of electronic commerce and the Internet.

10.7 Build an individual security solution

If each individual organization is unique, each security implementation is also unique. Once a security policy has been drawn up to suit the organization's needs and computing environment, building a security solution progresses logically. The items selected in previous stages determine the consistency of the deployment and the integration of the solution. The components are configured in accordance with the architectures and protocols defined, the flows to be secured, user profiles, applications and their location. Solutions may include:

- Firewall and VPN
- Intrusion Detection
- Anti-Virus
- Access Control and user Management

10.8 Implement a reference base and manage security globally

Once the various components have been installed, they have to be managed. Now, organizations have a wide range of security tools at their disposal, on their workstations, networks servers, and in their applications. The current issue is how to apply a coherent security policy. Coexistence with existing solutions must be allowed using a specific tool. This tool handles user access rights and computing resources, and simultaneously controls internal and remote connections. This global tool constitutes a common reference base, in compliance with the company's security policy.
10.9 **Train and inform users to raise their awareness**

Over and above the technological dimension, security is first and foremost a problem of human resources and organization. User involvement is fundamental. Motivation is vitally important, at all levels of the enterprise. On this point, security is similar to quality. In order to feel concerned, all the administrative and operating divisions in the organization must be involved.

10.10 **Audit regularly**

Finally, security must be monitored, audited, and watched closely on a daily basis. Security cannot be decreed. It must be dynamic, and it must become a reflex.

As a conclusion, making an enterprise information system secure implies a technological approach, but human resources and organization are just as important. Implementation requires well-structured security management!

### 11. THE FUTURE OF HACKING

Hacking in the future poses a great threat to internet. As more children grow up with computers, the new generation will know how to manipulate the digital world much more than the last. As internet grows, the potential to gain monetary rewards digitally will rise. Along with this higher potential will be an increase in the population of hackers.

By the actions taken in the 2002's, it is almost definite that hackers are creating more advanced programs every day. Hackers have extorted $400,000 from banks in New York and Europe, stolen $10 million from Citibank, and caused $800 million of losses in money and intellectual property of other corporations. From these incidents alone, it can be seen that the future is marked with bigger break-ins and larger monetary transfers.

Stealing credit card numbers and telephone access codes will become ubiquitous in the hacker community. Thus numbers will also become easily available to the public on the black market or underground. In addition to banks, crackers will attack mutual fund companies, stock brokerages, life insurance agencies, and corporate payroll systems. Web sites engaged in selling products and services will fall victim to break-ins where images and prices will be manipulated. A commercial website may hack a competitor’s website and make changes in order to gain a larger market share.

With increasing use of the internet to pay utility bills, make credit card payments, and do banking, more people will become dependent on computers for these tasks. This is similar to how people have become dependent on ATM's in the 2000's.Internet usage will increase day by day, but the amount of losses due to computer break-ins will also rise. As long as there is a way of making money on the internet by cracking systems, hackers will continue to grow in number. The level of sophistication in programs will also escalate. In the past the police have not been able to put an end to hacking. Although government agencies will attempt to stop break-ins in the future, they will not be successful.

### 12. CONCLUSION

This paper has covered the hacking history, hacker stages, hacker’s on Internet, hacking cultures, hacking technique used by different hackers, hacker ethics, etc. Hackers
have all engaged in illegal activities since the seventies. How to protect from hacker were discussed. In addition, how to protect from hackers was described. Use of Internet will increase, but hacking on the net will grow along with it.

Hackers’ ethics have ethical principles of the original Hacker Ethic suggest that it is the ethical duty of the hacker to remove barriers, liberate information, distribute power, recognize people based on their ability, and create things that are good and life enhancing through computers.

The new hacker ethic suggests that it is the ethical duty of new hackers to Protect information, Respect others privacy, Use the resources that are being wasted, Exceed unnecessary restrictions, Respect peoples' right to correspond with each other, Erase all the traces, Share the information, Be vigilant against cyber-tyranny, Always test security and system integrity of computer systems. Hacker cultures today come out of hobby hacking, academic hacking and network hacking. It is more or less based on an ethical code, interpreted and shared in different ways.

Hackers want to become elite. Everyone wants high status, and in hacker culture high status is measured by the amount of information one has. Since, in this culture, information is the key to status, and information is also a means to get power, an information economy has developed. Hackers trade information in hopes of getting more information and thus becoming more powerful. By understanding their development and motivations, it is hoped that elements leading to the development and building of motivations of hackers would be checked and controlled.

13. REFERENCES

http://www.securitymetrics.com/images/hackerchart.jpg
17. students.hamilton.edu.(2003). *The hackers and Hacking*.  
http://students.hamilton.edu/2006/msethi/100/oldhacker.html
http://www.ecob.iup.edu/jeb/December2001-issue/Beveren%20article2.pdf
19. webzone.k3.mah.se. (2003). *The hacker cultures*  
http://webzone.k3.mah.se/k3jolo/hackercultures/sources.htm
ON SELECTION OF AUTOREGRESSIVE ORDER IN CASE OF INCORRECTLY MODEL SPECIFICATION

Samir Safi
Department of Statistics, The Islamic University of Gaza, Gaza, Palestine
Email: samirsa@icloud.com

ABSTRACT

The purpose of this paper is to compare different autoregressive models performance in case of incorrectly model order determination. The autoregressive models are selected based on four important information criteria: the Estimated Mean Square Prediction Error (EMSPE), Akaike’s Information Criterion (AIC), Schwarz’s Bayesian Information Criterion (BIC), and the estimated white noise variance $\hat{\sigma}_w^2$. We generate different models with a second order autoregressive model, and consider the robustness of model selection based upon incorrect order. Combining a statistical interpretation of the four important information criteria and the principle of model parsimony, a model selection strategy is proposed. The main finding from the simulation is that over specification outperforms in selecting the most appropriate model, especially when the sample size is small compared with the number of parameters to be estimated.

KEYWORDS

Autoregressive; Information Criteria; AIC; BIC; EMSPE.

1. INTRODUCTION

The determination of an appropriate autoregressive model of order $p$ denoted by AR($p$) to represent an observed stationary time series involves a number of interrelated problems. These include the choice of the order of AR models and estimation of the mean, the model coefficients, and the white noise variance.

In time series model we are often faced with the problem of selecting appropriate values of the order of the model. It is not advantages from a forecasting point of view to choose the order of AR($p$) arbitrary large. Fitting a very high order model will generally result in a small estimated white noise variance, but when the fitted model is used for forecasting, the mean squared error of the forecasts will depend not only on the white noise variance of the fitted model but also on errors arising from estimation of the parameters of the model. These will be larger for higher-order models (Brockwell and Davis, 2002).

The probability to select the correct model increases approximately proportionally to the number of observations if all observations are performed under the same conditions, (Hoeven, N., 2005). An increase in observations will then go together with an increase in the number of model parameters. In this case, the power of the likelihood-ratio test will
increase with an increasing number of observations. However, the probability to choose the correct model with the Akaike's (1973) information criterion (AIC) will only increase if for each set of observations the Minimum Kullback–Leibler Divergence (MKLD) is more than 0.5. If the MKLD is less than 0.5, that probability will decrease. Wang and Liu (2006) showed that both AIC and Schwarz's (1978) Bayesian information criterion (BIC) are valid in selecting the most suitable stock–recruitment relationship. As far as the nested models are concerned, BIC is better than AIC.

Spriet and Herman (1983) showed that the error rate increases as the coefficient of the highest order time lag decreases to zero. Hossain et al. (2006) suggested that in most cases the forecasting performances of the models in question are quite satisfactory.

In this paper we have run an extensive simulation study to choose the best criteria by using different autoregressive models. We focus only on AR(p) process which are widely used in econometric studies. Four different information criteria were employed to select the best AR model out of the class of many possible models: the estimated mean square prediction error (EMSPE), Akaike's information criterion (AIC), Schwarz's Bayesian information criterion (BIC), and the estimated white noise variance $\hat{\sigma}_w^2$. The smaller these values, the better are the model selection.

The main finding from the simulation is that over specification outperforms in selecting the true model, especially when the sample size is small compared with the number of parameters to be estimated. In other words, the simulation suggests that selection model that would be dependent on the highest order time lag parameter could be superior. Furthermore, the simulation indicates the superiority of BIC over the other selected criteria.

This paper is organized as follows. In Section 2 we introduce the theoretical concepts of the model and the criteria model selection. Section 3 presents the complete simulation setup and results. Section 4 summarizes the results. Section 5 contains some concluding remarks.

2. AUTOREGRESSIVE MODEL AND CRITERIA MODEL SELECTION

In the classical regression model, the dependent variable is influenced by current values of the independent variables. In time series case, it is desirable to allow the dependent variable to be influenced by the past values of the independent variables and possibly by its own past values. If the present values can be modeled in terms of only the past values of the independent inputs, we have the enticing prospect that forecasting will be possible.

2.1 Autoregressive Model

For time series analysis, the general autoregressive model of order $p$, abbreviated AR(p), in terms of backward shift operator $B$ is of the form:

$$\left(1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p\right) X_t = W_t,$$  

(1)
where $X_t$ is a stationary time series process, $\phi_1, \phi_2, \ldots, \phi_p$ are constants ($\phi_p \neq 0$). Unless stated otherwise, $\{W_t; t = 0, \pm 1, \pm 2, \ldots\}$ is a Gaussian white noise series with mean zero and finite white noise variance $\sigma^2_w$. We introduce the definition of the variance of AR(2) process below.

**Definition:**

The variance of AR(2) process, $\left(1-\phi_1 B - \phi_2 B^2\right) X_t = W_t$ is given by

$$\sigma^2_X = \frac{\sigma^2_w}{\left(1 + \phi_2\right)\left((1-\phi_2)^2 - \phi_1^2\right)}$$

(2)

where $X_t$ is a stationary time series process, $\phi_1$ and $\phi_2$ are AR(2) coefficients, and $\sigma^2_w$ is white noise variance.

**2.2 Criteria Model Selection**

There are many criteria available in the literature for model selection, since the problem of model selection arises frequently in statistics. The statistical properties of those criteria are badly known. We introduce below a brief discussion of three different information criteria.

**2.2.1 EMSPE Criterion**

The estimated mean square prediction error, EMSPE, criterion was developed by Akaike (1969) to select the appropriate order of an AR process to fit to a time series $\{X_1, X_2, \ldots, X_n\}$. The idea is to choose the model $\{X_t\}$ to minimize the one-step squared error. The EMSPE is given by:

$$\text{EMSPE} = \sigma^2_w \left(\frac{n + p}{n - p}\right)$$

(3)

where $\sigma^2_w$ is the maximum likelihood estimate of the white noise variance of the AR(p) model.

**2.2.2 AIC Criterion**

A more generally applicable criterion for the model selection than EMSPE is the Akaike's Information Criterion (1973), known as the AIC.

The general Gaussian autoregressive moving average model, abbreviated ARMA(p,q), in terms of backward shift operator B is of the form:

$$\left(1-\phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p\right) X_t = \left(1 + \theta_1 B + \theta_2 B^2 + \cdots + \theta_q B^q\right) W_t$$

(4)

where $X_t$ is a stationary time series process, $\phi_1, \phi_2, \ldots, \phi_p$ and $\theta_1, \theta_2, \ldots, \theta_q$ are constants ($\phi_p \neq 0, \theta_q \neq 0$). Unless stated otherwise, $\{W_t; t = 0, \pm 1, \pm 2, \ldots\}$ is a Gaussian white noise series with mean zero and finite white noise variance $\sigma^2_w$. 
Suppose \( \{X_t\} \) is a Gaussian autoregressive ARMA(p,q) process with coefficient vector \( \Theta = (\phi, \theta) \) and white noise variance \( \sigma_W^2 \). For a zero-mean causal invertible ARMA(p,q) process, the AIC is given by:

\[
AIC(\Theta) = -2 \ln L_x(\Theta, S_x(\Theta)/n) + 2(p + q + 1)
\]  

(5)

where \( L_x(\Theta, S_x(\Theta)/n) \) is the likelihood function.

In other words, the AIC equals -2 times the log likelihood of the model plus 2 times the number of parameters fit. For fitting autoregressive models, Monte Carlo studies (Jones, 1975; Shibata, 1976) suggested that AIC has a tendency to overestimate \( p \).

### 2.2.3 BIC Criterion

Schwarz's Bayesian information criterion (1978), known as (BIC) is another criterion that attempts to correct the over fitting nature of the AIC. For a zero-mean causal invertible ARMA(p,q) process, the BIC is given by:

\[
BIC = (n - p - q) \log \left[ \frac{n \hat{\sigma}_W^2}{(n - p - q)^{-1}} \right] \\
+ n \left( 1 + \log \sqrt{2\pi} \right) + (p + q) \log \left[ \left( \frac{1}{n} \sum_{t=1}^{n} X_t^2 - n \hat{\sigma}_W^2 \right) (p + q)^{-1} \right]
\]  

(6)

where \( \hat{\sigma}_W^2 \) is the maximum likelihood estimate of the white noise variance.

As a rule of thumb, we would expect as small value as possible for all of these criteria to select the most appropriate autoregressive model.

### 3. SIMULATION STUDY

In this section, we consider the robustness of the four different information criteria for model selection such as EMSPE, AIC, BIC, and estimated white noise variance \( \hat{\sigma}_W^2 \).

The simulation examines the robustness of information criteria to model misspecification. In particular, how do information criteria perform when an AR(2) process is appropriate and we incorrectly assume the process is an AR(1)? We compare the four different information criteria to five different autoregressive models: the AR(2) model based on the correct disturbance model structures; the AR(1) incorrect disturbance model structures, i.e. when the order of the disturbance term is under estimated; and three more autoregressive models based on incorrect disturbance model structures with orders three, four and five denoted as AR(3), AR(4), and AR(5), i.e. when the order of the disturbance term is over estimated.

#### 3.1 Simulation Setup

Four finite sample sizes used are 20, 50, 200, and 500. We generated 1000 observations for each of the AR(2) error disturbances with five pairs of autoregressive coefficients; \((.2,-.9), (.2,-.7), (.2,-.1), (.8,-.9), \) and \((1.0,-.9)\). The parameter values of the disturbance AR(2) structures were chosen to provide different values of the variance of
the selected AR(2) processes. Table (1) shows the values of the variances of AR(2) processes, \( \sigma_X^2 \), for the selected values of AR(2) coefficients (Assuming that \( \sigma_w^2 = 1 \)).

Looking at Table (1) we see that the choices (1, -9), (.8, -9), and (.2, -9) give the highest variance among the selected AR(2) processes. However, the choices (.2, -1) and (.2, -7) give the smallest variance of the selected AR(2) processes.

**Table 1**

<table>
<thead>
<tr>
<th>(( \phi_1, \phi_2 ))</th>
<th>( \sigma_w^2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1.0, -9)</td>
<td>7.2797</td>
</tr>
<tr>
<td>(.8, -9)</td>
<td>6.3973</td>
</tr>
<tr>
<td>(.2, -9)</td>
<td>5.3221</td>
</tr>
<tr>
<td>(.2, -7)</td>
<td>1.9883</td>
</tr>
<tr>
<td>(.2, -1)</td>
<td>1.0446</td>
</tr>
</tbody>
</table>

We used the statistical software S-Plus to perform the simulation in order to compute the estimated values of the four information criteria. The following S-Plus functions are used to run the simulation and computations: `arima.sim` function to simulate a univariate AR(2) process, and `arima.mle` function to estimate ARIMA model by Gaussian maximum likelihood.

**3.2 Simulation Results**

Tables (2) through (6) show the complete simulation results of the four different information criteria for the model selection, EMSPE, AIC, BIC, and estimated white noise variance \( \hat{\sigma}_w^2 \). Each table presents the results for the four sample sizes considered, as well as all five selected autoregressive model of orders one through five. Each of the five pairs of AR(2) parametrizations is presented in a separate table.

We see that regardless of the sample size and selected pairs of AR(2) parametrizations, autoregressive model based on the correct disturbance model structures, AR(2), has the smallest values of the BIC and EMSPE criteria, and the estimated white noise variance \( \hat{\sigma}_w^2 \) is close to the true value of \( \sigma_w^2 = 1 \). For example, as shown in Table (2), when \( \phi = (.2, -9) \) with sample size \( T = 50 \), BIC = 103.721, EMSPE = 1.037 and \( \hat{\sigma}_w^2 = 0.958 \). In addition, there is insignificant difference between the AIC values for AR(2) and the other over estimated order of autoregressive models, i.e. AR with orders three and above.
Table 2:
Estimated Values of Information Criteria for \((\phi_1, \phi_2) = (.2,-.9)\)

<table>
<thead>
<tr>
<th>Size</th>
<th>Process</th>
<th>AIC</th>
<th>BIC</th>
<th>EMSPE</th>
<th>(\hat{\sigma}^2_w)</th>
</tr>
</thead>
<tbody>
<tr>
<td>20</td>
<td>AR(1)</td>
<td>82.472</td>
<td>65.331</td>
<td>5.746</td>
<td>5.199</td>
</tr>
<tr>
<td></td>
<td>AR(2)</td>
<td>51.823</td>
<td>43.803</td>
<td>1.085</td>
<td>0.888</td>
</tr>
<tr>
<td></td>
<td>AR(3)</td>
<td>49.783</td>
<td>45.730</td>
<td>1.119</td>
<td>0.827</td>
</tr>
<tr>
<td></td>
<td>AR(4)</td>
<td>47.819</td>
<td>47.385</td>
<td>1.149</td>
<td>0.766</td>
</tr>
<tr>
<td></td>
<td>AR(5)</td>
<td>45.788</td>
<td>48.725</td>
<td>1.161</td>
<td>0.697</td>
</tr>
<tr>
<td>50</td>
<td>AR(1)</td>
<td>215.273</td>
<td>170.582</td>
<td>5.466</td>
<td>5.252</td>
</tr>
<tr>
<td></td>
<td>AR(2)</td>
<td>137.075</td>
<td>103.721</td>
<td>1.037</td>
<td>0.958</td>
</tr>
<tr>
<td></td>
<td>AR(3)</td>
<td>135.258</td>
<td>106.973</td>
<td>1.057</td>
<td>0.938</td>
</tr>
<tr>
<td></td>
<td>AR(4)</td>
<td>133.235</td>
<td>109.678</td>
<td>1.072</td>
<td>0.913</td>
</tr>
<tr>
<td></td>
<td>AR(5)</td>
<td>131.427</td>
<td>112.333</td>
<td>1.090</td>
<td>0.892</td>
</tr>
<tr>
<td>200</td>
<td>AR(1)</td>
<td>890.945</td>
<td>711.343</td>
<td>5.368</td>
<td>5.315</td>
</tr>
<tr>
<td></td>
<td>AR(2)</td>
<td>563.804</td>
<td>395.726</td>
<td>1.014</td>
<td>0.994</td>
</tr>
<tr>
<td></td>
<td>AR(3)</td>
<td>562.018</td>
<td>400.574</td>
<td>1.020</td>
<td>0.990</td>
</tr>
<tr>
<td></td>
<td>AR(4)</td>
<td>560.249</td>
<td>405.097</td>
<td>1.025</td>
<td>0.985</td>
</tr>
<tr>
<td></td>
<td>AR(5)</td>
<td>558.613</td>
<td>409.497</td>
<td>1.031</td>
<td>0.981</td>
</tr>
<tr>
<td>500</td>
<td>AR(1)</td>
<td>2241.520</td>
<td>1787.211</td>
<td>5.320</td>
<td>5.299</td>
</tr>
<tr>
<td></td>
<td>AR(2)</td>
<td>1414.963</td>
<td>973.102</td>
<td>1.005</td>
<td>0.997</td>
</tr>
<tr>
<td></td>
<td>AR(3)</td>
<td>1,413.124</td>
<td>978.849</td>
<td>1.007</td>
<td>0.995</td>
</tr>
<tr>
<td></td>
<td>AR(4)</td>
<td>1,411.724</td>
<td>984.695</td>
<td>1.010</td>
<td>0.994</td>
</tr>
<tr>
<td></td>
<td>AR(5)</td>
<td>1,410.174</td>
<td>990.136</td>
<td>1.013</td>
<td>0.993</td>
</tr>
</tbody>
</table>

In addition, for the over estimated autoregressive models, the differences between the correct model, AR(2), and the over estimated models, i.e. order 3 and above, are statistically insignificant for all different information criteria. The simulation results show that AR(3) selection model performs better than the other over estimated models (Autoregressive models with orders four and five). For example, as shown in Table (3), when \(\phi = (.8,-.9)\) with sample size \(T = 200\), BIC = 395.545, 400.677, 405.449, and 410.057 for autoregressive models of orders 2, 3, 4, and 5.
Table 3:
Estimated Values of Information Criteria for \((\phi_1, \phi_2) = (.8,.9)\)

<table>
<thead>
<tr>
<th>Size</th>
<th>Process</th>
<th>(\phi_1, \phi_2)</th>
<th>AIC</th>
<th>BIC</th>
<th>EMSPE</th>
<th>(\hat{\sigma}_w^2)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>AR(1)</td>
<td>82.157</td>
<td>68.461</td>
<td>5.637</td>
<td>5.100</td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>AR(2)</td>
<td>51.728</td>
<td>44.186</td>
<td>1.087</td>
<td>0.890</td>
<td></td>
</tr>
<tr>
<td></td>
<td>AR(3)</td>
<td>49.943</td>
<td>46.537</td>
<td>1.133</td>
<td>0.838</td>
<td></td>
</tr>
<tr>
<td></td>
<td>AR(4)</td>
<td>47.914</td>
<td>48.329</td>
<td>1.156</td>
<td>0.771</td>
<td></td>
</tr>
<tr>
<td></td>
<td>AR(5)</td>
<td>45.739</td>
<td>49.713</td>
<td>1.165</td>
<td>0.699</td>
<td></td>
</tr>
<tr>
<td></td>
<td>AR(1)</td>
<td>214.915</td>
<td>174.659</td>
<td>5.448</td>
<td>5.235</td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>AR(2)</td>
<td>136.733</td>
<td>103.825</td>
<td>1.030</td>
<td>0.950</td>
<td></td>
</tr>
<tr>
<td></td>
<td>AR(3)</td>
<td>134.960</td>
<td>107.341</td>
<td>1.049</td>
<td>0.931</td>
<td></td>
</tr>
<tr>
<td></td>
<td>AR(4)</td>
<td>133.066</td>
<td>110.387</td>
<td>1.067</td>
<td>0.909</td>
<td></td>
</tr>
<tr>
<td></td>
<td>AR(5)</td>
<td>131.258</td>
<td>113.251</td>
<td>1.085</td>
<td>0.887</td>
<td></td>
</tr>
<tr>
<td></td>
<td>AR(1)</td>
<td>888.048</td>
<td>711.475</td>
<td>5.318</td>
<td>5.265</td>
<td></td>
</tr>
<tr>
<td>200</td>
<td>AR(2)</td>
<td>563.208</td>
<td>395.545</td>
<td>1.011</td>
<td>0.991</td>
<td></td>
</tr>
<tr>
<td></td>
<td>AR(3)</td>
<td>561.498</td>
<td>400.677</td>
<td>1.017</td>
<td>0.987</td>
<td></td>
</tr>
<tr>
<td></td>
<td>AR(4)</td>
<td>559.771</td>
<td>405.449</td>
<td>1.022</td>
<td>0.982</td>
<td></td>
</tr>
<tr>
<td></td>
<td>AR(5)</td>
<td>558.137</td>
<td>410.057</td>
<td>1.028</td>
<td>0.978</td>
<td></td>
</tr>
<tr>
<td></td>
<td>AR(1)</td>
<td>2235.191</td>
<td>1783.878</td>
<td>5.254</td>
<td>5.233</td>
<td></td>
</tr>
<tr>
<td>500</td>
<td>AR(2)</td>
<td>1413.489</td>
<td>972.048</td>
<td>1.002</td>
<td>0.994</td>
<td></td>
</tr>
<tr>
<td></td>
<td>AR(3)</td>
<td>1411.737</td>
<td>978.092</td>
<td>1.005</td>
<td>0.993</td>
<td></td>
</tr>
<tr>
<td></td>
<td>AR(4)</td>
<td>1410.302</td>
<td>984.111</td>
<td>1.007</td>
<td>0.991</td>
<td></td>
</tr>
<tr>
<td></td>
<td>AR(5)</td>
<td>1409.301</td>
<td>990.309</td>
<td>1.011</td>
<td>0.991</td>
<td></td>
</tr>
</tbody>
</table>

For the other selected autoregressive models, mainly when the order of the autoregressive model is under estimated, i.e. AR(1), we see that regardless of the sample size, the information criteria AIC, BIC, and EMSPE attain the largest value among the other selected autoregressive models. In addition, the estimated white noise variance \(\hat{\sigma}_w^2\) is overestimated. For example, as shown in Table (4), when \(\phi=(1.0, -.9)\) with sample size \(T = 50\), AIC = 215.366, BIC = 175.716, EMSPE = 5.428, and \(\hat{\sigma}_w^2 = 5.215\). This result shows that AR(1) model performs poorly.
Table 4:  
Estimated Values of Information Criteria for $(\phi_1, \phi_2) = (1.0,-.9)$

<table>
<thead>
<tr>
<th>Size</th>
<th>Process</th>
<th>AIC</th>
<th>BIC</th>
<th>EMSPE</th>
<th>$\hat{\sigma}_W^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>20</td>
<td>AR(1)</td>
<td>83.006</td>
<td>69.882</td>
<td>5.965</td>
<td>5.397</td>
</tr>
<tr>
<td></td>
<td>AR(2)</td>
<td>51.837</td>
<td>44.710</td>
<td>1.085</td>
<td>0.888</td>
</tr>
<tr>
<td></td>
<td>AR(3)</td>
<td>49.801</td>
<td>47.052</td>
<td>1.121</td>
<td>0.829</td>
</tr>
<tr>
<td></td>
<td>AR(4)</td>
<td>47.810</td>
<td>49.080</td>
<td>1.152</td>
<td>0.768</td>
</tr>
<tr>
<td></td>
<td>AR(5)</td>
<td>45.748</td>
<td>50.758</td>
<td>1.168</td>
<td>0.701</td>
</tr>
<tr>
<td>50</td>
<td>AR(1)</td>
<td>215.366</td>
<td>175.716</td>
<td>5.428</td>
<td>5.215</td>
</tr>
<tr>
<td></td>
<td>AR(2)</td>
<td>137.518</td>
<td>104.959</td>
<td>1.046</td>
<td>0.965</td>
</tr>
<tr>
<td></td>
<td>AR(3)</td>
<td>135.605</td>
<td>108.508</td>
<td>1.064</td>
<td>0.944</td>
</tr>
<tr>
<td></td>
<td>AR(4)</td>
<td>133.675</td>
<td>111.689</td>
<td>1.081</td>
<td>0.921</td>
</tr>
<tr>
<td></td>
<td>AR(5)</td>
<td>131.977</td>
<td>114.835</td>
<td>1.102</td>
<td>0.902</td>
</tr>
<tr>
<td>200</td>
<td>AR(1)</td>
<td>887.235</td>
<td>711.230</td>
<td>5.294</td>
<td>5.241</td>
</tr>
<tr>
<td></td>
<td>AR(2)</td>
<td>561.936</td>
<td>394.573</td>
<td>1.005</td>
<td>0.985</td>
</tr>
<tr>
<td></td>
<td>AR(3)</td>
<td>560.402</td>
<td>400.030</td>
<td>1.011</td>
<td>0.982</td>
</tr>
<tr>
<td></td>
<td>AR(4)</td>
<td>558.690</td>
<td>404.966</td>
<td>1.017</td>
<td>0.977</td>
</tr>
<tr>
<td></td>
<td>AR(5)</td>
<td>557.216</td>
<td>409.882</td>
<td>1.024</td>
<td>0.974</td>
</tr>
<tr>
<td>500</td>
<td>AR(1)</td>
<td>2240.675</td>
<td>1789.945</td>
<td>5.316</td>
<td>5.294</td>
</tr>
<tr>
<td></td>
<td>AR(2)</td>
<td>1415.533</td>
<td>974.419</td>
<td>1.007</td>
<td>0.999</td>
</tr>
<tr>
<td></td>
<td>AR(3)</td>
<td>1414.121</td>
<td>980.966</td>
<td>1.010</td>
<td>0.997</td>
</tr>
<tr>
<td></td>
<td>AR(4)</td>
<td>1412.475</td>
<td>986.937</td>
<td>1.012</td>
<td>0.996</td>
</tr>
<tr>
<td></td>
<td>AR(5)</td>
<td>1411.822</td>
<td>993.646</td>
<td>1.016</td>
<td>0.996</td>
</tr>
</tbody>
</table>

Furthermore, as the sample size increases, AR(1) selection model performs much worse. For example, as shown in Table (5), for $\phi = (0.2,-.7)$, when $T=20$, BIC = 48.027 and 40.761 for AR(1) and AR(2), respectively, i.e. BIC for AR(1) is 1.18 times its corresponding criterion BIC for AR(2). Similarly, for the other selected samples sizes $T=50$, 200, and 500, the information criterion BIC for AR(1) are 1.26, 1.31, 1.34 times its corresponding criterion for AR(2), respectively.
Table 5:
Estimated Values of Information Criteria for \((\phi_1, \phi_2) = (0.2, -0.7)\)

<table>
<thead>
<tr>
<th>Size</th>
<th>Process</th>
<th>((\phi_1, \phi_2) = (0.2, -0.7))</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>AIC</td>
<td>BIC</td>
</tr>
<tr>
<td>20</td>
<td>AR(1)</td>
<td>65.680</td>
</tr>
<tr>
<td></td>
<td>AR(2)</td>
<td>51.528</td>
</tr>
<tr>
<td></td>
<td>AR(3)</td>
<td>49.706</td>
</tr>
<tr>
<td></td>
<td>AR(4)</td>
<td>47.804</td>
</tr>
<tr>
<td></td>
<td>AR(5)</td>
<td>45.760</td>
</tr>
<tr>
<td>50</td>
<td>AR(1)</td>
<td>171.461</td>
</tr>
<tr>
<td></td>
<td>AR(2)</td>
<td>137.476</td>
</tr>
<tr>
<td></td>
<td>AR(3)</td>
<td>135.597</td>
</tr>
<tr>
<td></td>
<td>AR(4)</td>
<td>133.718</td>
</tr>
<tr>
<td></td>
<td>AR(5)</td>
<td>131.981</td>
</tr>
<tr>
<td>200</td>
<td>AR(1)</td>
<td>695.059</td>
</tr>
<tr>
<td></td>
<td>AR(2)</td>
<td>562.754</td>
</tr>
<tr>
<td></td>
<td>AR(3)</td>
<td>560.841</td>
</tr>
<tr>
<td></td>
<td>AR(4)</td>
<td>559.107</td>
</tr>
<tr>
<td></td>
<td>AR(5)</td>
<td>557.275</td>
</tr>
<tr>
<td>500</td>
<td>AR(1)</td>
<td>1751.516</td>
</tr>
<tr>
<td></td>
<td>AR(2)</td>
<td>1414.785</td>
</tr>
<tr>
<td></td>
<td>AR(3)</td>
<td>1412.958</td>
</tr>
<tr>
<td></td>
<td>AR(4)</td>
<td>1411.125</td>
</tr>
<tr>
<td></td>
<td>AR(5)</td>
<td>1409.298</td>
</tr>
</tbody>
</table>

In addition, the simulation results show that AIC attains its smallest value for high order of autoregressive models; in other words, AIC has a tendency to over estimate the order of autoregressive models. On the other hand, BIC corrects the over fitting of the AIC. For example, as shown in Table (5), for \(\phi = (0.2, -0.7)\), when T=50, AIC = 137.476, 135.597, 133.718, and 131.981, for autoregressive models of orders 2 through 5, respectively. For the other criterion, BIC = 101.337, 103.190, 104.737, and 106.184 for autoregressive models of orders 2 through 5, respectively. The results for the other pairs of autoregressive coefficients and sample sizes mimic the same behavior of this chosen example.

Finally, AR(1) performs nearly as efficiently as the other selected models for all selected sample sizes when \(\phi = (0.2, -0.1)\). For example, as shown in Table (6), when \(\phi=(0.2, -0.1)\) with sample size T = 20, AIC = 54.469 and 52.371, BIC = 37.807 and 37.384, EMSPE = 1.083 and 1.115, and \(\hat{\sigma}_W^2 = 0.980 \) and 0.912 for AR(1) and AR(2), respectively. This result is not surprising since the choice of \(\phi=(0.2, -0.1)\) gives the smallest value of the variance of the process, \(\sigma_X^2\), as shown in Table (1). Furthermore, the choice of \(\varphi_2 = -0.1\) indicates that the serially correlated disturbance very nearly AR(1) since \(\varphi_2\) is close to zero.
4. DISCUSSION

In investigating the simulation results in the previous section, we observe the following significant results. First and foremost we notice that regardless of the sample size and selected pairs of AR(2) parametrizations, autoregressive model based on the correct disturbance model structures, AR(2), has the smallest values of the BIC and EMSPE criteria, and the estimated white noise variance $\hat{\sigma}_W^2$ is close to the true selected value of $\sigma_W^2 = 1$. For the over estimated autoregressive models, the differences between the correct model, AR(2), and the over estimated models, i.e. order 3 and above, are not statistically significant for all different information criteria. The simulation results show that AR(3) selection model performs better than the other over estimated models.

For the other selected autoregressive models, mainly when the order of the autoregressive model is under estimated, i.e. AR(1), we see that regardless of the sample size, the information criteria AIC, BIC, and EMSPE attain the largest value among the other selected autoregressive models. In addition, the estimated white noise variance $\hat{\sigma}_W^2$ is overestimated.

Table 6:
Estimated Values of Information criteria for $(\phi_1, \phi_2) = (.2, -.1)$

<table>
<thead>
<tr>
<th>Size</th>
<th>Process</th>
<th>AIC</th>
<th>BIC</th>
<th>EMSPE</th>
<th>$\hat{\sigma}_W^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>20</td>
<td>AR(1)</td>
<td>54.469</td>
<td>37.807</td>
<td>1.083</td>
<td>0.980</td>
</tr>
<tr>
<td></td>
<td>AR(2)</td>
<td>52.371</td>
<td>37.384</td>
<td>1.115</td>
<td>0.912</td>
</tr>
<tr>
<td></td>
<td>AR(3)</td>
<td>50.539</td>
<td>37.145</td>
<td>1.163</td>
<td>0.859</td>
</tr>
<tr>
<td></td>
<td>AR(4)</td>
<td>48.539</td>
<td>36.664</td>
<td>1.195</td>
<td>0.797</td>
</tr>
<tr>
<td></td>
<td>AR(5)</td>
<td>46.516</td>
<td>36.544</td>
<td>1.215</td>
<td>0.729</td>
</tr>
<tr>
<td>50</td>
<td>AR(1)</td>
<td>140.070</td>
<td>96.589</td>
<td>1.041</td>
<td>1.000</td>
</tr>
<tr>
<td></td>
<td>AR(2)</td>
<td>137.530</td>
<td>96.027</td>
<td>1.046</td>
<td>0.965</td>
</tr>
<tr>
<td></td>
<td>AR(3)</td>
<td>135.669</td>
<td>95.914</td>
<td>1.065</td>
<td>0.944</td>
</tr>
<tr>
<td></td>
<td>AR(4)</td>
<td>133.753</td>
<td>95.884</td>
<td>1.082</td>
<td>0.922</td>
</tr>
<tr>
<td></td>
<td>AR(5)</td>
<td>131.937</td>
<td>95.769</td>
<td>1.101</td>
<td>0.901</td>
</tr>
<tr>
<td>200</td>
<td>AR(1)</td>
<td>566.994</td>
<td>386.711</td>
<td>1.017</td>
<td>1.007</td>
</tr>
<tr>
<td></td>
<td>AR(2)</td>
<td>562.981</td>
<td>385.842</td>
<td>1.011</td>
<td>0.991</td>
</tr>
<tr>
<td></td>
<td>AR(3)</td>
<td>561.183</td>
<td>386.490</td>
<td>1.016</td>
<td>0.986</td>
</tr>
<tr>
<td></td>
<td>AR(4)</td>
<td>559.425</td>
<td>386.942</td>
<td>1.021</td>
<td>0.981</td>
</tr>
<tr>
<td></td>
<td>AR(5)</td>
<td>557.610</td>
<td>387.225</td>
<td>1.026</td>
<td>0.976</td>
</tr>
<tr>
<td>500</td>
<td>AR(1)</td>
<td>1418.860</td>
<td>963.987</td>
<td>1.008</td>
<td>1.004</td>
</tr>
<tr>
<td></td>
<td>AR(2)</td>
<td>1412.149</td>
<td>961.118</td>
<td>1.000</td>
<td>0.992</td>
</tr>
<tr>
<td></td>
<td>AR(3)</td>
<td>1410.357</td>
<td>962.455</td>
<td>1.002</td>
<td>0.990</td>
</tr>
<tr>
<td></td>
<td>AR(4)</td>
<td>1408.570</td>
<td>963.540</td>
<td>1.004</td>
<td>0.988</td>
</tr>
<tr>
<td></td>
<td>AR(5)</td>
<td>1406.829</td>
<td>964.465</td>
<td>1.006</td>
<td>0.986</td>
</tr>
</tbody>
</table>
Furthermore, as the sample size increases, AR(1) selection model performs much worse. Finally, AR(1) performs nearly as efficiently as the other selected models for all selected sample sizes when $\phi = (.2, -.1)$ since the choice of $\phi_2 = -.1$ indicates that the serially correlated disturbance very nearly AR(1) since $\phi_2$ is close to zero.

5. CONCLUDING REMARKS

This paper has investigated an important statistical problem concerning selection of the order of autoregressive models in the presence of auto-correlated disturbances. In particular, we have discussed the autoregressive model selection based on four important information criteria; EMSPE, AIC, BIC, and the estimated white noise variance $\hat{\sigma}_w^2$.

The simulation shows that over specification outperforms in selecting the true model, especially when the sample size is small compared with the number of parameters to be estimated. In addition, the simulation indicates the superiority of BIC over the other selected criteria. In addition, the simulation suggests that selection model that would be dependent on the highest order time lag parameter could be superior. Furthermore, BIC criteria corrects the over fitting of the AIC.

REFERENCES

NEW VARIATIONAL MODEL FOR IMAGES ELECTIVE SEGMENTATION

Haider Ali, Noor Badshah and Ghulam Murtaza
Department of Basic Sciences, University of Engineering and Technology, Peshawar, Pakistan.
Email: haider_uop99@yahoo.com; noor2knoor@gmail.com; murtaza_uop99@yahoo.com

ABSTRACT

In image segmentation, images with fuzzy edges, low frequencies, intersecting regions of almost homogeneous intensities and images with intensity en-homogeneity are the challenging images for existing segmentation models [1, 2, 3, 4, 7, 16]. Recently, we proposed a new selective segmentation model [1] which works efficiently for many images except the above mentioned challenging images. Based on the concept of covariance and local statistical functional [3], we propose a new model of image selective segmentation. Experimental results ensure that the performance of our newly proposed model is far better than the previous works in segmenting such tough images.

KEYWORDS:
Segmentation; Covariance (CoV); Level Set; Functional Minimization; Total Variation (TV).

1. INTRODUCTION

Image segmentation is an important branch of image processing. Segmentation of images means to divide an image into its constituent parts which are homogeneous in some sense like intensity or texture etc. Various models have been developed for image segmentation tasks and few examples are, watershed algorithms [10], region growing and emerging [15] and Mumford-Shah functional minimization [4]. Let $z(x, y)$ be a given image defined on a rectangular domain $\Omega$.

Mumford and Shah (MS) proposed general model:

$$
\min_{u, \Gamma} F(u, \Gamma) = \mu \cdot \text{length}(\Gamma) + \lambda \int_{\Omega} |z - u|^2 dx + \int_{\partial \Omega \setminus \Gamma} |\nabla u|^2 dx
$$

to automatically detect edge $\Gamma$ of and piecewise smooth version $u(x, y)$ of $z(x, y)$.

The Chan-Vese (CV) [2] proposed the following model:

$$
F(c_1, c_2, \Gamma) = \mu \cdot \text{length}(\Gamma) + \lambda_1 \int_{\text{inside}(\Gamma)} |z - c_1|^2 dx dy + \lambda_2 \int_{\text{outside}(\Gamma)} |z - c_2|^2 dx dy,
$$

where $z(x, y)$ is a given image, $c_1$ and $c_2$ are constants denoting average values of $z(x, y)$ inside and outside of $\Gamma$ respectively. The CV model can be interpreted as a
special case of the piecewise constant MS model when restricted to two phases. Although above segmentation models are useful for various segmentation tasks, but sometimes we are not interested in the whole image but interested in only one part of the image, this process is classify as image selective segmentation. Recently, we proposed a new model of selective segmentation [1], which is discussed in next section. In contrast with existing models, there we ensured the best efficiency of our model in noisy images and best performance in terms of robustness and accuracy. In images with low contrast, intersecting regions with homogeneous intensities and in images having un-illuminated objects, our that model may fails to work. Now we will equip our model with a new type of image data fidelity term that can work better even when a given image has overlapping regions with almost homogeneous intensities or when edges of a given image are not prominent. This data fidelity term is based on the concept of covariance. Our experimental results validate the excellent performance of this new type of fidelity term based model, in contrast with the old model.

We organize this paper in the following way. In section 2 we give a review of the Badshah-Chen model [1]. In section 3 we introduce our proposed new model of minimization and the corresponding Euler-Lagrange equation. In Section 4 we include an additive operator splitting (AOS) method for solving the PDE. In section 5 we include experimental results.

2. THE BADSHAH-CHEN MODEL (BC)

To segment a given image \( z(x, y) \), The Badshah and Chen Model [1] is

\[
\min_{c_1, c_2, \Gamma} F(\Gamma, c_1, c_2)
\]

where

\[
F(\Gamma, c_1, c_2) = \mu \int_{\Gamma} d(x, y) g(|\nabla z|) \, ds \\
+ \lambda_1 \int_{\text{outside}(\Gamma)} (z - c_1)^2 \, dxdy + \lambda_2 \int_{\text{inside}(\Gamma)} (z - c_2)^2 \, dxdy
\]  

(1)

where \( \mu, \lambda_1, \lambda_2 \) are positive constants and are used for assigning different weights, \( c_1 \) and \( c_2 \) are the average intensities outside and inside a contour \( \Gamma \) respectively.

The function \( d(x, y) \) is a distance function defined in [7] as:

\[
d(x, y) = \prod_{i=1}^{n} \left( 1 - e^{-\frac{(x - x_i)^2}{2\sigma^2}} e^{-\frac{(y - y_i)^2}{2\sigma^2}} \right), \quad \forall (x, y) \in \Omega,
\]

where the marker set

\[A = \{(x_i, y_i) : i = 1, 2, 3, ..., m\}\]

are the given geometrical constraints and we want that the boundary of an object of interest to be detected. Practically, the user needs to only click on the point of interest in the image to setup A.
is denoting edge detector function and the following one is a popular choice.

\[ g(|\nabla z|) = \frac{1}{1 + |\nabla z|^2} \]

In level set formulation, equation (1) becomes,

\[ F(\phi, c_1, c_2) = \mu \int_{\Omega} d(x, y)g(|\nabla z|)\delta(\phi)|\nabla \phi|dx\,dy + \lambda_1 \int_{\Omega} |z(x, y) - c_1|^2 H(\phi)dx\,dy \\
+ \lambda_2 \int_{\Omega} |z(x, y) - c_2|^2 (1 - H(\phi))dx\,dy, \]

where

\[ H(x) = \begin{cases} 
1 & \text{if } x \geq 0 \\
0 & \text{if } x < 0
\end{cases} \quad \text{and} \quad \delta(x) = H'(x) \]

Since the Heaviside function is not differentiable at the origin, a regularized version of Heaviside function is used [2, 5, 6, 14].

\[ H_\varepsilon(w) = \frac{1}{2} \left( 1 + \frac{2}{\pi} \arctan \left( \frac{w}{\varepsilon} \right) \right), \quad \delta_\varepsilon(w) = \frac{\varepsilon}{\pi(\varepsilon^2 + w^2)}. \]

The regularized functional \( F_\varepsilon(\phi, c_1, c_2) \), is given by

\[ F_\varepsilon(\phi, c_1, c_2) = \mu \int_{\Omega} d(x, y)g(|\nabla z|)\delta_\varepsilon(\phi)|\nabla \phi|dx\,dy + \lambda_1 \int_{\Omega} |z(x, y) - c_1|^2 H_\varepsilon(\phi)dx\,dy \\
+ \lambda_2 \int_{\Omega} |z(x, y) - c_2|^2 (1 - H_\varepsilon(\phi))dx\,dy. \]

Keeping \( \phi \) fixed and minimizing \( F_\varepsilon(\phi, c_1, c_2) \) with respect to \( c_1 \) and \( c_2 \), we have

\[ c_1(\phi) = \frac{\int_{\Omega} z(x, y)H_\varepsilon(\phi)dx\,dy}{\int_{\Omega} H_\varepsilon(\phi)dx\,dy}, \quad c_2(\phi) = \frac{\int_{\Omega} z(x, y)(1 - H_\varepsilon(\phi))dx\,dy}{\int_{\Omega}(1 - H_\varepsilon(\phi))dx\,dy}, \]

assuming that the curve has a non-empty interior and non-empty exterior in \( \Omega \).

Keeping \( c_1 \) and \( c_2 \) fixed and minimizing \( F_\varepsilon \) with respect to \( \phi \) yields the following Euler-Lagrange equation for \( \phi \):

\[
\begin{cases}
\delta_\varepsilon(\phi) \left[ \mu \text{div} \left( G(x, y) \frac{\nabla \phi}{|\nabla \phi|} \right) 
- \lambda_1(z(x, y) - c_1)^2 + \lambda_2(z(x, y) - c_2)^2 \right] = 0 & \text{in } \Omega, \\
\frac{G(x, y)\delta_\varepsilon(\phi) \partial \phi}{|\nabla \phi| \partial n} = 0 & \text{on } \partial \Omega,
\end{cases}
\]

where

\[ G(x, y) = d(x, y)g(|\nabla z|), \]
\( \mathbf{n} \) is the unit exterior normal to the boundary \( \partial \Omega \), and \( \frac{\partial \phi}{\partial \mathbf{n}} \) is the normal derivative of \( \phi \) at the boundary.

The above PDE may be considered as a steady state form of the following evolution equation:

\[
\frac{\partial \phi}{\partial t} = \delta_{\epsilon}(\phi) \left[ \mu \nabla \cdot \left( G(x, y) \frac{\nabla \phi}{|\nabla \phi|} \right) \right] \\
- \lambda_1 (z - c_1)^2 + \lambda_2 (z - c_2)^2, \quad \text{in } \Omega \\
\phi(t, x, y) = \phi_0(x, y), \quad \text{in } \Omega.
\]

A balloon term \([7], \alpha G(x, y) |\nabla \phi|\), was added for robustness and iteration initialization, where \( \alpha \) is a constant. To solve the above evolution equation, the Additive Operator Splitting Method (AOS) was used \([9, 11]\). Since the function

\[
\int_{\Gamma} d(x, y) g \left( |\nabla z| \right) ds
\]

similar to \([7, 13]\), is the first term of Badshah-Chen model.

The purpose was to minimize their proposed functional so that to find the unknown boundary curve \( \Gamma \).

Noise in an image produces trouble for an edge detector function in detecting edges. As this model depends on edge detector function so it usually fails to work in noisy images and in images with discrete or fuzzy edges. Isotropic Gaussian smoothing helps in smoothing \( z \), but unfortunately it also smooth the edges.

The second term of BC model is the image data fidelity term

\[
\int_{\text{outside}(\Gamma)} (z - c_1)^2 dxdy + \int_{\text{inside}(\Gamma)} (z - c_2)^2 dxdy
\]

which belongs to the CV model proposed in \([2]\).

This term empowered the BC model to work in noisy images and to obtain robustness in terms of number of iteration and so in this way the advantages of the CV model are utilized in BC model.

However there are images which are challenging for segmentation tasks. In particular, MRI and CT images with fuzzy edges, unilluminated organs and overlapping homogeneous regions.

Since the BC model \((1)\) involves the fidelity term or region detector.

\[
\int_{\text{outside}(\Gamma)} (z - c_1)^2 dxdy + \int_{\text{inside}(\Gamma)} (z - c_2)^2 dxdy
\]

This fidelity term carries its advantages as well as its weaknesses to BC model.
Experimental results revealed that BC model does not work well, while working with such challenging images due to detection of spurious objects. For segmenting such tough and challenging images, better region detectors are required.

3. THE PROPOSED MODEL

In this section we will develop a new variational model for image selective segmentation, and for this, we will observe test results obtained using our three new energy functional. Based on the performance, we will select one of these three newly developed models and will compare our model with the latest existing models [1,7]. Since a fidelity term usually carries a given image information only and many models can be seen in the literature with such fidelity terms [1, 2, 3]. The idea of covariance can help to develop an energy functional which involves statistical information of a given image as well as it incorporates guidance from an enhanced version of that image.

Since covariance of two variables $X$ and $Y$, is given by:

$$ C = \frac{1}{n-1} \sum (X - \bar{X})(Y - \bar{Y}), $$

or

$$ C = \frac{1}{n-1} \sum (XY - \bar{X}\bar{Y}) $$

(3)

Along with a given image $u_0(x, y)$, we wish to use as an enhanced version of $u_0(x, y)$, the averaging convolution image $u^*(x, y)$, i.e., $u^*(x, y) = g_k * u_0(x, y)$ with $g_k$ an averaging convolution operator of window size $k \times k$.

The equation (3) can help to construct a fidelity term in continuous settings that can include, a given image information and incorporates information of an enhanced version of the image, in the main model.

Thus we construct a fidelity term by using $L^2$-norm and therefore the product of means is replaced by a constant say $c_i$ by observing equation (2).

We get:

$$ \int_{\text{outside}(\Gamma')} (uu^* - c_1)^2 dx dy + \int_{\text{inside}(\Gamma')} (uu^* - c_2)^2 dx dy. $$

(4)

So we obtained our first new model by combining the fidelity term given in (2) with function given in (4) as follows:

$$ F_1(\Gamma, c_1, c_2) = \mu \int_{\Gamma} d(x, y) g(|\nabla z|) ds $$

$$ + \lambda_1 \int_{\text{outside}(\Gamma')} (uu^* - c_1)^2 dx dy + \lambda_2 \int_{\text{inside}(\Gamma')} (uu^* - c_2)^2 dx dy. $$

The figures 1 and 2 display the performance of $F_1$ model. The simple test 1(b) shows
Fig.1: A simple example of segmenting a hardware image in which $F_1$ completed the assigned task. Size = $256 \times 256$.

Fig.2: Segmenting a real brain image using $F_1$ (a) original image with initial contour; (b) the uncompleted task by $F_1$. Size = $256 \times 256$.

Although, the model $F_1$ is equipped with double images region information, but still it fails to work in figure 2 because of its global behavior. It missed valuable local information and failed to detect the desired object/region. Next, we proceed towards our second model.

Usually in region-based models two terms namely, global term and local term, contribute mainly. A global term helps in detecting the main structure formed by objects/regions in an image, whereas, a local term helps in capturing small and valuable details.

For utilizing local information, we use local fitting term proposed in [3] given by,

$$
\int_{\text{outside} (\Gamma)} \left( (g_k \ast z - z) - d_1 \right)^2 dxdy + \int_{\text{inside} (\Gamma)} \left( (g_k \ast z - z) - d_2 \right)^2 dxdy,
$$

where $d_1$ and $d_2$ are the average intensities of the difference image $g_k \ast z - z$ inside and outside $\Gamma$, respectively and $z = uu^*$. By denoting the difference image $g_k \ast z - z$ with
$z^*$, we obtained our second new model by combining the function given in (2) with fitting term given in (5) as follows:

$$F_2(\Gamma, d_1, d_2) = \mu \int_\Gamma d(x, y) g(|\nabla z|) ds + \lambda_1 \int_{\text{outside}(\Gamma)} (z^* - d_1)^2 dxdy + \lambda_2 \int_{\text{inside}(\Gamma)} (z^* - d_2)^2 dxdy.$$

The figures 3 and 4 show the performance of $F_2$. The simple test 3(b) exhibits that

**Fig.3:** A simple example of segmenting a hardware image in which $F_2$ completed the assigned task. Size = 256 × 256.

**Fig.4:** Segmenting a real brain image using $F_2$ (a) original image with initial contour; (b) the uncompleted task by $F_2$. Size = 256 × 256.

The figures 3 and 4 show the performance of $F_2$. The simple test fig: 3(b) exhibits that the $F_2$ model detected the desired object/region but the second test fig: 4(b) clearly displays that the $F_2$ model fails to detect accurately, the desired portion in the given brain image.
Furthermore, figure 4(b) is describing, that the local information is alone not sufficient, for accurate detection both global and local information are necessary.

To develop our third new model, we combine the fitting terms given in (4) and (5) with the function given in (2) as follows:

\[
F_3(\Gamma, c_1, c_2, d_1, d_2) = \mu \int_{\Gamma} d(x, y) g(|\nabla z|) ds \\
+ \int_{\text{outside}(\Gamma)} \left[ \lambda_1 (uu^* - c_1)^2 + \lambda_2 (z^* - d_1)^2 \right] dx dy \\
+ \int_{\text{inside}(\Gamma)} \left[ \lambda_1 (uu^* - c_2)^2 + \lambda_2 (z^* - d_2)^2 \right] dx dy.
\]

Fig.5: A simple example of segmenting a hardware image in which \( F_3 \) completed the assigned task. Size = 256 \( \times \) 256.

Fig.6: Segmenting a real brain image using \( F_3 \) model. (a) original image with initial contour; (b) the successful result of \( F_3 \). Size = 256 \( \times \) 256.

The figures 5 and 6 exhibit the performance of \( F_3 \). The simple test 5 validates that the \( F_3 \) model detected the desired object/region. Similarly, the second test fig: 6 also shows that the \( F_3 \) model successfully detected accurately, the desired portion in the given brain image. Since the \( F_3 \) model involves data fidelity terms that incorporates both local and global information. Therefore, for the comparison with the latest existing
selective segmentation models [1, 7], we prefer to use the $F_3$ model in next experimental section. For simplicity, we give briefly only the minimization of $F_3$ here.

Denoting $F_3$ by $F$ in level set formulation is given by:

$$F(\phi, c_1, c_2, d_1, d_2) = \mu \int_{\Omega} d(x, y) g(|\nabla z|) |\delta(\phi)| \nabla \phi | dx dy$$

$$+ \int_{\Omega} \left[ \lambda_1 (z - c_1)^2 + \lambda_2 (z^* - d_1)^2 \right] H(\phi) dx dy$$

$$+ \int_{\Omega} \left[ \lambda_1 (z - c_2)^2 + \lambda_2 (z^* - d_2)^2 \right] (1 - H(\phi)) dx dy,$$

By considering the following regularized minimization problem

$$\min_{\phi, c_1, c_2, d_1, d_2} F_c(\phi, c_1, c_2, d_1, d_2),$$

where

$$F_c(\phi, c_1, c_2, d_1, d_2) = \mu \int_{\Omega} d(x, y) g(|\nabla z|) |\delta_c(\phi)| \nabla \phi | dx dy$$

$$+ \int_{\Omega} \left[ \lambda_1 (z - c_1)^2 + \lambda_2 (z^* - d_1)^2 \right] H_c(\phi) dx dy$$

$$+ \int_{\Omega} \left[ \lambda_1 (z - c_2)^2 + \lambda_2 (z^* - d_2)^2 \right] (1 - H_c(\phi)) dx dy,$$

minimization of $F_c(\phi, c_1, c_2, d_1, d_2)$ with respect to $c_1, c_2, d_1, d_2$ and $\phi$ leads to the following solutions:

$$c_1(\phi) = \frac{\int_{\Omega} z(x, y) H_c(\phi) dx dy}{\int_{\Omega} H_c(\phi) dx dy}, \quad c_2(\phi) = \frac{\int_{\Omega} z(x, y) (1 - H_c(\phi)) dx dy}{\int_{\Omega} (1 - H_c(\phi)) dx dy},$$

$$d_1(\phi) = \frac{\int_{\Omega} z^*(x, y) H_c(\phi) dx dy}{\int_{\Omega} H_c(\phi) dx dy}, \quad d_2(\phi) = \frac{\int_{\Omega} z^*(x, y) (1 - H_c(\phi)) dx dy}{\int_{\Omega} (1 - H_c(\phi)) dx dy},$$

and $\phi$:

$$\delta_c(\phi) \left[ \mu \text{div} \left( \frac{\nabla G(x, y)}{\nabla \phi} \right) \right]$$

$$+ \lambda_1 \left( (z - c_1)^2 + (z - c_2)^2 \right) + \lambda_2 \left( (z^* - d_1)^2 + (z^* - d_2)^2 \right) = 0 \quad \text{in} \ \Omega,$$

$$\frac{\partial \phi}{\partial n} = 0 \quad \text{on} \ \partial \Omega,$$

The above PDE may be considered as a steady state form of the following evolution equation.
By implementing, AOS method to solve the above PDE (6), the following system of equations is obtained.

\[
\frac{\partial \phi}{\partial t} = \delta_\epsilon(\phi) \left[ \mu \nabla \cdot \left( G(x, y) \frac{\nabla \phi}{\nabla \phi} \right) - \lambda_1 (z - c_1)^2 \right. \\
\left. - \lambda_2 (z^* - d_1)^2 + \lambda_2 (z - c_2)^2 + \lambda_2 (z^* - d_2)^2 \right], \quad \text{in } \Omega
\]

\[
\phi(x, y, t) = \phi_0(x, y, 0), \quad \text{in } \Omega.
\]

(6)

By implementing, AOS method to solve the above PDE (6), the following system of equations is obtained.

\[
(I - 2\Delta t A_l(\Phi^p))\Phi^{p+1}_l = f^p, \quad \text{for } l = 1, 2,
\]

and \[\Phi^{p+1}_l = \frac{1}{2} \sum_{i=1}^{2} \Phi^{p+1}_l,\]

where I is the identity matrix and \( A_l \) for \( l = 1, 2 \) a tridiagonal matrix.

**EXPERIMENTAL RESULTS**

In this section we give some simulation results.

For convenience, we shall denote by

- M-1 --- the Gout model
- M-2 --- the BC model
- M-3 --- the proposed CLM.

Below we give comparison results of these three methods.

In both M-2 and M-3, we use \( \lambda_1 = \lambda_2 = \lambda \).

The behavior of M-2 can be seen in figures fig: 7(c) and fig: 8(c).

On the other hand the following experiments reveal the unsatisfactory results of M-1 model. In figures fig: 7(b) and fig: 8(b) it can be seen that M-1 did not completed the tasks. In contrast with M-1, M-2, the best performance of M-3 can be seen from the experiments. The experiments also exhibit that M-3 is best in accurate and fast detection and successful in the images in which these two models are unable to work. In figures fig: 7(d) and fig: 8(d) the successful detection by M-3 can be easily seen.
5. COMMENTS AND CONCLUSION

In summary, it has been observed from the experiments that, while performing selective segmentation on the challenging images having nearly equal intensity regions or fuzzy edges, the performance of M-2 is less effective, as in such cases it is often observed that the active contour crosses the boundary of an object of interest in the image and therefore the existing model is unable in detecting the actual boundary and consequently the region of interest in the image. In contrast, the proposed new M-3 outperforms all existing methods.

Thus a new image selective segmentation model is proposed. Based on covariance and local statistical functional, this model has the shown best results in comparison with the latest works [1, 7]. The above tests ensure that this model segments those images accurately in which the latest models are unable to perform segmentation.
REFERENCES

A CONCEPT FOR PERSONAGE’S LEARNING STYLE CENTERED TECHNOLOGY BASED ADAPTIVE PEDAGOGICAL SYSTEM

Muhammad Anwaar Saeed* and Khalil Ahmed
School of Computer Science, National College of Business Administration and Economics, Lahore Pakistan
Email: *m_a_saeed@yahoo.com

ABSTRACT

Every person has his own learning style and preferences. In our existing educational systems, one is bound to study an offered subject even if it is mismatched with one’s learning style. Currently, same method is used to assess all students of a course, because it is difficult to consider an individual’s learning style in existing conventional educational systems. With the increasing involvement of technology in education, one can design a course capable of adapting an individual’s or group’s learning style and preferences. In this work, a concept of such pedagogical system is outlined. Students will learn more in such pedagogical system and a better assessment of students will also be possible.

INTRODUCTION

Joyce Yukawa (Yukawa, 2006) has observed the importance of reflection and co-reflection concept in an online course by a comparative case study. It is observed that co-reflection was central to both individual and group cognition for the participants of case study. It is also observed through case study results that co-reflection can be used as core process for group orientation. Though during the case study same environment/resources were provided, students have used them differently (Yukawa, 2006).

Knowledge in computer science field grow rapidly with new innovations which leads to the frequent curriculum revision. Also multicultural class is getting a norm in educational institutes nowadays. Alaoutinen et al. (Alaoutinen, Heikkinen, and Porras, 2010) have proposed a framework for computer science field considering five dimensions, namely pedagogical framework, pedagogical pattern, curriculum, learning outcomes, and teaching methods. The interaction between these dimensions is also considered. Supporters of collaborative learning believe that exchange of ideas among group increases both interest and critical thinking of participants. Humans gather and process the information in different ways, called learning styles and according to Dunn learning styles can be compared to fingerprints that can change during the years. It is observed that cognitive abilities depend on the surroundings in which we grow up and spend most of our time. One’s learning style takes longer to change if style preferences are stronger. There is no general agreement to classify different existing learning styles. For example, Kolb’s Experiential Learning Theory has two dimensions, the Myers-Briggs Type Indicator has four dimensions, and Felder & Silverman has presented five dimensions. It is observed that the consideration of learning style preferences for the selection of teaching method and learning objective has a positive effect on learning. To calculate/determine the learning styles of participants, Alaoutinen et al. (Alaoutinen, Heikkinen, and Porras, 2010) have used Felder-Silverman Learning Style Model in their
study. It is observed during study that to achieve best learning outcomes a learner should be taught using his/her preferred learning style. It is also observed that level of knowledge and skills may vary among group members. In general, participating students were satisfied with the course (Alaoutinen, Heikkinen, and Porras, 2010).

Both individual learning environment and collaborative workspace is possible using internet. A wrongly selected may have a negative impact on outcomes in collaborative learning. There are only few systems developed in an attempt to adapt an individual’s learning style. But there is still confusion about the aspect of learning style worth modeling and what can be done differently for learners with different learning styles. Using Felder & Silverman Model, we can classify any individual into a selected style in each of five dimensions i.e. active/reflection, sensing/intuitive, visual/verbal, sequential/global, and inductive/deductive. In 1980, a new area of study based on different theories, Computer-Supported Collaborative Learning (CSCL) is emerged. Group formation is an important aspect in collaborative learning. Different studies show that for specific tasks homogeneous groups perform better but for tasks with broader range heterogeneous groups give better results. It is observed that if students are allowed they arrange themselves in homogeneous groups. In order to group students based on features, two approaches exist. In one approach, for a given students working on comparable problems, we need to find pair of students who can benefits from cooperation in joint session. Whereas in second approach, for a given group we need to select/generate a problem that forms an adequate challenge for the group as a whole. Results of case study show that correlation between any two dimensions is small in all cases. (Alfonseca et al. Sep 2006)

Each student has his/her own learning style and he/she learn easily if taught by considering his/her learning style. Many developed adaptive systems have focus to provide learning experience matching with learning style, but these systems ignore situation where student has to learn against his learning style. Studies show that students with different learning styles behave differently in non-adaptive courses. It is observed that learners with strong learning style preferences had low performance in comparison to learners with no strong learning style preferences. It is also observed that reflective learners have performed better scores than active learners in a mismatched course, which show that adaptivity is more important for active learners. (Kinshuk, Liu, and Graf, 2009)

The way a person likes to process information and deal with task is called its intellectual style which encompasses cognitive, learning, and thinking styles. It is observed from the literature that abilities generally grow as a function of age. Also men tend to demonstrate higher levels of spatial abilities than women. Whereas, females are tend to be more concrete than male. Males have stronger preference for abstract sequential learning style; on the other hand females have stronger preference for abstract random style. Children are more field-dependent and their field independence increase till their adulthood. Adult learners are mostly field-independent; afterward their field independence decreases and people in older age are mostly field-dependent. The preferred ways of using one’s abilities are considered thinking styles. Sternberg has categorized 13 thinking styles that fall along the dimensions of function, form, level, scope, and learning. Zhang and Sternberg have re-conceptualized their 13 styles into three types. First type of thinking styles tend to be more creativity-generating and they denote higher level of cognitive complexity. Second type of thinking styles suggest norm-favoring tendency and they denote lower level of cognitive complexity. Third type of thinking styles may manifest the characteristics of the styles from first and
second type depending on the stylistic demand of a specific task. Research suggest that thinking styles vary as a function of both personal (e.g. age and gender) and environmental (e.g. nature of academic discipline) characteristics. It is also observed that thinking styles have an effect on student’s school performance and teacher’s teaching almost in every part of world. A significant change in students’ performance is observed when thinking styles are considered during teaching. It is observed from the case study results that older students and male students scored significantly higher on both analytical and creative abilities. It is also observed that female and older students tended to score significantly lower on first type of thinking styles. It is observed that age and gender do make a difference in the relationship between thinking styles and abilities (Zhang, 2010).

Cavaliar et al. (Cavaliar, and Klein, 1998) have conducted a multivariate analysis of variance for performance, attitude, and time on task. It is observed from the results that subjects with instructional objectives performed significantly. It is also observed that individuals spent significant portion of time on instruction and practice than subjects working in cooperative dyads. It is also observed that cooperative dyads that received objectives have shown more helping behavior and on-task group behavior. And dyads that did not receive orienting activities have shown significant on-task individual behavior and off-task behavior (Cavaliar, and Klein, 1998).

It is observed from different previous educational research work that in collaborative activities students learn most of the taught material effectively and retain this knowledge for longer period in contrast to the conventional teaching. Kyprianidou et al. (Kyprianidou et al. 2011) have used Roundsepp Problem Solving Styles Inventory to identify the learning styles of students and then used this information in construction of hybrid groups. In general, collaboration has a positive effect but in some cases students feel uncomfortable in collaborative activities and too it as a threat to their performance. Psychological factors based diversity among team members is recommended by many researchers for team building. It is observed through research work that size of a successful group is limited to five at most, in general. The rationale of the study is that learning style based group formation has benefits for student collaboration at many levels. Both quantitative and qualitative data indicates that students were happy with learning style idea and mostly agreed with their identified profile. Results show that learning styles were prominent during collaboration. With the awareness of strengths and weaknesses of each others, students began to honor differences and consider everyone as talented and competent (Kyprianidou et al. 2011).

Learning outcomes of e-learning students is affected by instructor feedback and student learning style. If teaching styles are modified to accommodate each student’s learning style then quality of education is enhanced significantly. While designing a lesson plan, teacher needs to add activities to reflect different learning styles. Teachers using various instructional strategies are proved to be effective than teachers using single strategy. Liaw and Huang suggested four elements for consideration while developing e-learning environments i.e. environmental characteristics, environmental satisfaction, learning activities, and learners’ characteristics. In adaptive educational system, learning style profile of a student is managed and used to adapt the presentation and navigation of instructional content. Development of an adaptive e-learning system needs appropriate selection of learning style model and instrument, course content consistent with various
learning styles, and determination of level and degree of adaptation of domain content (Markovic, and Jovanovic, 2011).

It is observed from the results of concept map analysis that structural changes in students’ knowledge were rare. No obvious correlation between the quantitative and qualitative measures of student achievement is observed. Concept mapping can be used to document personal understanding as this method is realistic during university teaching. It is observed that rote or non-learning is occurred where prior knowledge is poor. Concept mapping can be used to identify the students at risk of failure. Prior knowledge has a significant role in quality of student learning (Hay, Wells, and Kinchin, 2008).

Currently universities are compelled to adopt technology for broadening the access to education, personalizing the teaching experience, and specialized skill. With increasing use of learning technologies, teachers and students have more chances to identify and explore suitable teaching and learning style. Distance learning is becoming easy with new mediums to transfer information and teacher-student interaction. Blended learning is defined by integrating e-learning with traditional learning programs and technology. Kolb has classified the learning styles in four areas: concrete experience, reflective observation, abstract conceptualization, and active experimentation. Learning method of each learning style is different from others. Using Kolb model, learning styles are categorized as assimilators, accommodators, divergents, and convergers. Learning styles have a significant impact in the development and use of blended learning media. It is observed through the case study results that students have a positive opinion about blended learning method. It also appears from the results that students preferred blended learning to only face to face teaching (Uger, Akkoyunlu, and Kurbanoglu, 2011).

Cultural style based instruction planning or repertoires culturally-based experiences are two contrasting ways to response cultural differences in a diverse university classroom. Generally, learning styles are used to address the individuals’ learning methods differences (Maccleave, and Eghan, 2010).

**ADAPTIVE PEDAGOGICAL SYSTEM CONCEPT**

One can conclude from the previous work on pedagogical systems that in current systems courses and teaching methods are not developed/adapted according to individual’s/group’s learning style. Though in conventional systems this adaptive nature is hard to embed due to limitations of system by nature, but technology dependent educational systems has the potential to provide such adaptability. Such adjustable educational system can fulfill the needs of both individuals and groups.

In the proposed technology based system, we need to arrange semester activities in accordance to the learning style preferences of an individual/group. For this purpose, we need to have lecture of a topic delivered with different teaching methods and similarly course contents like example and practical activities etc. This means that we need to have a collection of multiple versions of same semester activities of a single course matching with different learning styles. Individual’s learning style is calculated before the commencement of class and based on this contents are picked from the pool and allocated. For a homogeneous group selection of contents is an easy task, but for a heterogeneous group a prior calculation is required to select proper set of contents and teaching method. As this activity is performed for each course, it will cater any change in
learning style. Also contents can be updated at anytime, even if a specific learning styles/teaching method needs reformation.

Following figure No. 1 portrays the important activities needs to be performed in an adaptive pedagogical system. As discussed, adaptive nature cannot be produced in as educational systems until we have course contents mapped to all nature of learning style preferences. Therefore multiple versions course contents are required and need to be developed when designing a course. After the contents development, delivery of these contents needs to be considered as one can retain knowledge for longer period if proper teaching method is adopted. Therefore a suitable teaching method is selected for each version of course content. At third stage, we select a proper set of semester activities for an individual/group in accordance with learning styles. Different modules of the proposed pedagogical system are shown in figure 2. This model can be adopted in existing pedagogical systems by removing the teaching selection method only.
A new concept for a technology based pedagogical system is proposed in this work. By applying the proposed concept, we can consider individuals’ / Groups’ preferences to ease their learning activities. Due to its adaptive nature, this proposed concept of pedagogical system has the flexibility to adjust contents and teaching methods according to preferences. It will also enhance the quality and knowledge level of learners.

FUTURE WORK

In future, implementation details of different modules will be defined and a working model of this concept will be constructed.

REFERENCES

COGNITIVE MODEL FOR SOCIAL BOTS TO LEARN OPPO NENTS MIXED STRATEGY USING ADAPTIVE NEURAL NETWORKS

Asma Kanwal¹ and Wajahat M. Qazi¹,²
¹ CIRL, Department of Computer Science, GC University, Lahore, Pakistan. Email: asmakhan93@gmail.com
² School of Computer Science, National College of Business Administration and Economics, Lahore, Pakistan

ABSTRACT

Since the paradigm shift from behavioral to motivational systems, the need for the development of social robots has become challenging. In this new emerging scenario, both man and machine have their own motives and goals, which instinctively drive them to maximize their payoffs with reference to game theory. In such conditions, the aspect of common knowledge of rationality may not be applicable as highlighted in various past studies. On the other hand, a single general-purpose strategy learning system may not also be appropriate as many individuals may act differently in different situations. These considerations encourage a profile-based strategic learning mechanism to be devised in machines that could enable it to collaborate with humans in a social structure where human strategies may change randomly and the rules of the game are known. This study reports a cognitive model for machine that could enable them to learn profile-based mixed strategies of its opponent using adaptive neural networks.

KEYWORDS

Artificial Neural Networks, Game Theory, Machine Learning.

1. INTRODUCTION

Since the engineering of intelligent robotics, scientists were designing behavior-based robots that perform their actions on the basis of predefined rules [1]. Indeed these systems have the ability to carry out pre-programmed tasks through learning but lack of the element of motivation that enable them to define new goals and tasks [2]. Social robots are motivation-based robots, which can set their motivation and change their objective during their social interaction with human in any environment.

In game theory any human can never show identical playing strategy almost in its every game. A human can make changes in his/her strategy by keeping in view the current playing strategy of opponent. Humans understand the opponent strategies in game from one’s own perspective and from others perspective to attain equilibrium. According to Game theory any person understands opponent’s playing strategy on the basis of self-assessment. Self-assessment is obtained when the rules of game, rationality of player and player’s pay off functions are in common knowledge of players. Certain cases exists in which such inclination is not define [3,4]. Furthermore, shift in behavior from non-equilibrium to equilibrium does not support the introspective theories [3]. The efficient
approach to determine the behavior of the opponent through learning playing strategies [5]. Learning provides ability in any player to attain equilibrium [6] and prevents the game from being complex [7].

In any game such artificial agent requires having learning ability to understand the playing strategy of human [8]. Learning is a process of acquiring skill or knowledge [9], whereas specific knowledge of a person and personal observation assist any human to expect others goals [10]. Behavioral imitation provides ability in human for understanding other human behavior in any game [11]. In order to cover the observation perspective of any artificial agent it requires grasping the implicit knowledge from the environment. Implicit learning is to find out the regularities in the environment and then expresses them explicitly [12]. In any game, if a machine predicts the playing strategy of opponents then it can generate its moves towards opponents in better way.

Many methods are described previously but evolutionary neural network (ENN) with its adaptive learning characteristic is considered as a most appropriate technique to provide learning ability in any artificial agent. Decision making in such games is based on player observation on series of temporal events and their learned knowledge. Considering this fact there is a need of designing such motivation-based social robot, with adaptive learning ability to play Paper-Scissor-Rock (PSR) game with opponents in any social environment [13], based on Tapped Delay Line Neural Network (TDNN) [14].

This study reports a theoretical cognitive model for machine based on TDNN [14] has ability to implicitly learn the playing strategy of any opponents through its observation and can predict any opponent’s next move in currently playing game on the basis of its current observation and past experience.

2. GAME THEORY

Game is any interaction between agents that govern by a set of rules specifying the possible moves for each participant and a set of outcomes for each possible combination of moves [15]. Socially interactive game’s agent requires property to attain equilibrium in the game through their self-assessment in any particular situation. According to Heap, when two rational players have the same information, they must draw the same inference and conclusion [16].

2.1 Problem with Multiple Equilibrium

A conceptual problem is; there are multiple equilibrium, for in the absence of an explanation of how players come to expect the same equilibrium, their game need not correspond to any equilibrium at all [3]. Harsanyi and Selten in 1988 make an idea that players coordinate their expectations using a common selection procedure. They do not explain how such a procedure comes to common knowledge [3]. Hypothesis of exact common knowledge of payoffs and rationality can not to be apply to every game [3,4]. The change in state from non-equilibrium to equilibrium in game is difficult to resolve with just introspective theories. They give the concept of incorporating learning theories in games [3].
2.2 Common Knowledge of Rationality

Common knowledge of rationality (CKR) is a procedure through which any game acquires equilibrium in it. In CKR, every agent knows the move of other agent before the agent plays its move [16]. Sometimes player chooses move randomly on some probabilistic bases. This randomization can become helpful for the player to become unpredictable for the opponent. The concatenated form of random and planned moves of any player delineates mixed strategy for the opponent [16]. A way through which any player can understand the mixed strategy of the opponent and attain equilibrium in the game is Nash Equilibrium. Nash equilibrium makes a confirmation about the expectations of each player for opponent choice. Nash equilibrium support Consistent Alignment of Beliefs (CAB) in CKR. In Nash equilibrium, no player can play better only by changing its own strategy; it required some confirmation on the prediction of opponent move [16]. Using CAB, if a player knows his opponent plans, he/she get a confirmation about his/her choice and will not change his/her belief, as well as, if opponent knows his/her plan, he/she does not require to change his/her belief. Means it is going to confirm its choice by considering the opponent choice prior then opponent play its move [16].

3. MACHINE LEARNING

For any social robots, social interaction, it requires to provide ability in the machine to do learning from the environment.

3.1 Machine in the Learning Process

Learning from the environment consist on a learner and its an environment. Where, learner is an active agent, which has some decided and changeable features in it. It can use these characteristics to achieve its goals, through performing actions in the environment. Whereas environment is a passive agent, which can change its internal states according to the actions, perform by active agent on it or by any other source [17]. Different types of learning techniques are available; supervised learning, unsupervised learning, reinforcement learning. Tapped Delay Line Neural Network based on reinforcement learning. Reinforcement learning is a process in which observer do the assessment of training that either network is doing ‘good’ or ‘bad’ response to the training patterns [18]. It is a technique in which past good actions are collected to generate an effective strategy [8,19].

3.2 Implicit and Explicit Learning and Knowledge

The observation perspective of any model is acquired through grasping the implicit knowledge from the environment. Implicit knowledge is the beliefs, perspectives and mental models of any one’s mind that generates from the expertise, intuitions or insight of the person. This is a knowledge type which is not explicitly defined [20]. Implicit knowledge is grasping through the implicit learning. Implicit learning produces such knowledge about which any one does not know anything [12]. Implicit learning defines the reasoning behind certain action and then these reasoning can generate explicit rules. The knowledge extracted from implicit learning is explicit knowledge. Explicit knowledge gets more elaborative with the extension in the implicit knowledge [21].
4. COGNITIVE MODEL SOCIAL INTERACTION (CMSI)

CMSI is a cognitive model for social bots to socially interact with human in PSR game and learn their playing mix strategy. In game, artificial agent can learn and understand the playing strategy of its opponent through two techniques. One is, to observe its opponent playing tactic during a game. Second is to understand opponent playing policy through his/her past, in addition of current experience, which can get by playing same game multiple times with its opponent. This learning can improve the responses of artificial agent towards its opponent as it starts to play game according to its opponent’s moves. Not every human playing strategy is fixed, so artificial agent try to change its playing strategy in any game, through predicting or understanding the moves of human.

![CMSI Model](image)

**Fig. 1: CMSI Model**

CMSI model consist on different cognitive modules. Figure 1 shows the CMSI model for social robots. Further, this study gives a functional description of different cognitive modules of CMSI model.

*Sensory Memory*: In game playing social robots firstly through its software sensors get initial necessary information about opponent and rules of game. Sensory memory is a storage area, temporarily store sensed information and understand some basic information about opponent.

*Perception Evaluation*: This module based on previously sensed information checks that artificial agent has any past playing experience with that human. If opponent is unknown and playing first time then profile information is store in his/her memory for future games. In other case it will recall that human memorized profile for validity of perception.

*Working Memory*: Working memory works as buffer for internal activities and provides temporary storage. In working memory based on output generated by perception evaluation module, agent performs identification of interacting person. If it identifies that
human then it recalls its previous learned implicit knowledge from long-term memory, to currently choose its moves in game in better way through predicting playing mix strategy of human. In other case, if artificial agent plays game first time with human then it will start observing current playing tactics of human and predict human next move in current game. This module is also responsible to check if during game it senses any new person entry then how to manage both interactions.

**Long Term Memory (LTM):** It is permanent memory of social robot. It is responsible to provide any memorized learned implicit knowledge of human after identification process. The last part which any agent does in its observation is, to memorize this predicted implicit knowledge in its memory for future game with that person.

**Social Response:** Predicted implicit knowledge will help any social robot to decide its move towards other human. The more accurate prediction will generate responses that are more approved and help agent to attain equilibrium state in game.

5. **CONCLUSION**

This study reports a cognitive model, which is under development in C# language. CMSI model shows that social robots, based on this model, will play rationally. When artificial agent play its game on rational basis and has learning ability that it can through learned implicit knowledge predict the next move of opponent in current game then there are some possibilities exist that any artificial player may attain an equilibrium state in its playing.

**REFERENCES**


SKANS: SECURE KNOTS AT THE ENDS

Syed Ali Raza, Tanwir Ahmad and Asma Basharat
Department of Computer Science, GC University, Lahore, Pakistan
Email: arjafri104@gmail.com

ABSTRACT

Cloud computing is emerging as a sizzling topic in the field of information technology but along with this rapid growth in the popularity of cloud computing the cloud community is facing real danger when it comes to security. This issue becomes severely critical in case of data storage as for commercial organizations security of data is of prime significance. The aim of this research paper is to propose a framework through which the data storage in cloud computing may be made more secure and trustworthy.

1. INTRODUCTION

Cloud computing is a computing scenario which provides consumption based computing facilities to users on demand [1]. The essence of cloud computing is to provide everything as a service. According to [2] a cloud may provide computational power, software’s, platform and data storage as a service. For executing computational operations the importance of data is such that computing without the data is as uncommon as data without the computing. This association of data and computing power makes huge tasks to be workable.

Most of the time first thing organizations or individual users demand is secure data storage and in cloud computing this demand is increasing exponentially. In the race of providing efficient and cheap data storage facility normally the cloud vendors ignores this demand of secure data storage or treat it as a least significant matter [2]. For commercial organizations or governments secure data storage is more critical issue while using cloud computing. Such organizations or governments cannot afford leakage of data in any case. To provide services to such organizations cloud vendors should have to think seriously about security.

In case where cloud security depends on web based portals then breaking this sort of security is not a big deal due to man in the middle attack and due to the fact that administrators or the technical staff of the cloud servers may creep into users confidential data. Claim of this study is that to maintain the overall security in cloud computing, the persistence of access policies and reliability of data storage deserves vigilant consideration. Focus of this paper is towards secure data storage on cloud servers. By securing the ends of communication and developing a secure mechanism of communication this target can easily be achieved. Here some risk factors have been discussed which have prime significance while discussing about cloud security.

A. Risk factors in Unsecured Cloud Computing

Risk is the probability that the system’s asset will be damaged or abused by the threats that exploit the vulnerabilities. Cloud computing faces several kind of risks. First
of these risks is legal exposure which can cause severe financial damage, and can even be a life-ending event for an organization. As many governments don’t allow vendors to provide these kind of services because governments, financial and healthcare institutions have particular security and privacy concerns. [5] Customer Confidence is also a risk factor due to the fact that customers will churn if they feel a company can’t secure their data so there should be a standardized security mechanism so that users may feel secure using cloud services. The customer translates bad data practices as lack of respect and thus may lose confidence in any organization that can’t keep information safe. [5] Loss, unavailability or misuse of assets is a prime concern in case of cloud computing. User’s data stored on cloud servers is the most important asset which if lost or modified without users consent can damage the cloud vendors market worth.

2. MOTIVATION

In the Information Technology (IT) arena, a threat is anything that would hurt company’s assets, or anything else contained on the computer network as well as the systems themselves. Basically cloud users face three major types of attacks each of which has been briefly described below.

A. Threats associated with Transactions

These threats are based on the attacks during transactions of data. If the cloud servers rely on the traditional web forms security then automated brute force attack can create hazardous problems. In many cases (e.g. Facebook, eyeOS) if the user id is known then attacker can easily brute force the password of the account.

Another kind of transactional attack is the Man in the Middle (MITM). MITM can create severe problems for any communication but in the case of cloud computing it becomes more critical. Through MITM important data can be modified by the attacker where an attacker can insert himself ‘between’ two communicating parties and intercept the traffic, read it, and perhaps alter it. [8] MITM attack can also be used to capture packets during communication.

In replay attack, attacker captures the packet(s) of secure traffic during the authenticated session and later he/she replays the captured traffic to cause unpredicted results. Attack is successful if the source cannot differentiate between original and duplicate traffic. Replay attack can cause a severe damage during data transfer, as replay attack can be combined with other attacks to cause more harm [18].

Web applications sometimes depend on different browsers and in case browsers start malfunctioning then it can harm overall security mechanism. Un-patched or older versions of IE contain several vulnerabilities that can lead to memory corruption, spoofing and remote execution of unreliable scripts or code without user permission. Similarly to IE, unpatched or older versions of firefox also hold multiple vulnerabilities [9].

B. Threats associated with Remote Data Storage

The crucial demand for cloud computing is to deliver secured storage. Few types of threat related with Cloud Data Storage have been illustrated here. In Confidentiality attack, attackers try to read information from a storage system without proper authorization like sniffing storage traffic. If an attacker achieves administrator level access to a system, they can typically explore storage with few safeguards and can result
in the loss of critical confidential business data on the Internet. The other risk is that damaging information can be posted about an organization that can negatively affect its status.

**Integrity attacks** attempt to modify information in a storage system without appropriate authorization. Modification may include creating, changing, appending, writing, and deleting both data and metadata. Integrity attacks include
- Modifying of Metadata
- Modifying Transaction Command

**C. Internal attack**

Threats originating from within the network are called internal attacks. For examples include malicious employees, employees that are not malicious but make mistakes. The administrators or help desk technician with too many rights to the system getting distress about management’s decision to not give them another raise this year, can become a potential threat [7]. She/he may corrupt his/her personal user share data that has not yet been backed up on the server that was vital to the company or company’s customer [7]. These days most of data storage organizations think they are mainly at threat from external attack, when they should be looking nearer to home. As Peter Wood said in [6] “many firms are spending too much time and resources securing themselves against external attacks, and ignoring the threat from insiders who can be tapped up by organized crime. Criminal groups are aware that the insider route can be the easier one to compromise systems, especially as many organizations have spent considerable resources to secure themselves against outside attacks”

### 3. SECURITY CONCERNS

Professor Adi Shamir said in [2] that “we could be facing a real danger that hackers would be able to take one of these datacenters out of commission - and that would have a catastrophic effect”. This means that the security parameters defined for cloud computing are not much efficient till the moment although a lot of work is being done to make cloud services secure and trustworthy. Gartner [3] illustrated that Cloud computing is filled with hazardous security risks and hence pointed out seven specific security issues for consideration of potential Cloud users. Intel information technology presents its concern about security of cloud computing in [4]. According to it once the data is stored on cloud server the owner of data typically doesn’t aware of the location of data. There is quite a chance that the owner’s data reside at same resource where its competitor’s data is kept. Secondly it states that the security standards are lacking. It also insists on managing service level agreement.

### 4. EXISTING SCENARIOS

The reason is that some cloud vendors are using the traditional security mechanism like simple account password for authentication and confidentiality which is so vulnerable against the above mentioned attacks. On the other hand some big vendors have their own security mechanisms but they are also not strong enough to defeat the attackers such that according to [11] “privacy and security on Google cloud services have many weaknesses, including the fact that data on Google's servers is not protected with strong encryption.”
SKANS: Secure Knots at the Ends

eyeOS [13] is a free cloud operating system that uses SSL for secure authentication and VFS (Virtual File System) for data storage. eyeOS security mostly rely on access policies. Access policies can protect your data from other the users but it cannot protect data from those personnel who manage and control access policies and from internal attacks. The security model designed for a largest cloud storage provider Amazon as given in [16] can be summarized as Amazon provides its users access key ID and a secret access Key. For sign in request user have to perform a hash of the request using secret access key then attach the hash with request and forward it to Amazon which verifies the hash and if verified process the request. This security model has been evaluated in [12] and according to this evaluation the security model presented by Amazon is just like a simple password that can be reset by anyone who could receive emails at the registered email address. The main problem in the existing mechanism being used by many cloud vendors is that they are somehow dependent on the traditional web browsers and so inherit some basic flaws. According to [1] cloud normally rely on web forms using SSL to manage their personal account information and it allows users to change or receive their private keys through emails which is totally unsafe due to unencrypted communication.

Using SSL, Kerberos [15] and other third party authentication mechanism increase latency during client and server authentication. Authentication Server (AS) can also create bottleneck situation and may become additional points of attack for attacker. In [11] George Reese have mentioned some rules for particularly Amazon’s security but these rules can also be applied generally to other cloud applications. One thing which lacks in the Reese’s rules is that he doesn’t speak about the mechanism through which these rules can be achieved. In this paper, some of his rules have been used to establish our framework for securing the data stored on cloud servers. One major work which is in progress for securing cloud services is the Cloud Security Alliance (CSA) [14]. In this paper different domains have been established to enhance the overall performance of the cloud. Although CSA is doing great efforts but the aim of this alliance is towards interoperability rather than securing the data.

5. SKANS

SKANS has two different levels. In the first level security against the internal attacks has been proposed. Internal attacks can be stopped by applying access policies. The access policies will be defined to the user at the time of registration in terms of service level agreement. These policies would base on the following rules.

1. Data would be kept encrypted on the cloud server using user’s public key.
2. Administrator or technical staff cannot move, update or delete any record from user’s data.
3. Whenever any administrator logs in to any user’s data storage whatever actions they perform would be logged.
4. Log of each action performed by administrators or technical staff will be sent to users via email.
5. In the process of registration the user’s IP address will be saved in a Trusted IP List (TIPL) maintained by the server.
6. Users can only login to the server from their trusted IP addresses.
7. If user will try to login using any IP address not listed in the server then the user will be asked additional secret questions for TIPL authentication and the IP will be added in TIPL.

8. On three consecutive wrong answers of the secret questions during TIPL authentication that IP address will be blacklisted. No request will be entertained from blacklisted IP.

The second level of SKANS consists on SKANS Application (SAPP) and Counter Threat Mechanism (CTM). In the beginning when a client registers itself on cloud server, a couple of keys (each key of 128 bits) will be generated (private and public key). Public key \((PU_U)^1\) of client will be stored on the server while the private key \((PR_U)\) will only be kept by the user and there will be no record of private key on server. SAPP will be downloaded at the client side. SAPP act as an interface between client and Cloud server. SAPP consists on a communication mechanism as shown in figure 1, an interface for interaction with the server to access the data storage account and Random Number Generator (RNG). SKANS uses an algorithm proposed in [17] as RNG. It has period over \(2^{128} - 1\) and \(2^{256} - 1\) on 32bit and 64bit CPU platform respectively. It has some advantageous characteristics like no statistical weaknesses, satisfactory randomness and higher speed.

When users sign in the SAPP, RNG will then generate 128 bit Random Number (RN) known as Nonce which will be encrypted by the client’s private key using AES. After that Encrypted RN (ERN) will be encrypted again with client’s private key using RSA and sent to the server. Server will decrypt the ERN using client’s public key and encrypt the ERN with the public key and send back to the client.

**Algorithm 1: SKANS Authentication**

**Client side:**
1. \(R \leftarrow \text{RNG}\)
2. \(R \leftarrow \text{Encrypt}_{AES}(R, PR_U)\)
3. \(R_E \leftarrow \text{Encrypt}_{RSA}(R, PR_U)\)
4. Send \([R_E + ID_U]\) to Server

**Server Side:**
5. \(R_S \leftarrow \text{Decrypt}_{RSA}(R_E, PU_U)\)
6. \(RS_S \leftarrow \text{Encrypt}_{RSA}(R_S, PU_U)\)
7. Send \([RS_S + ID_S]\) to Client

**Client side Continue:**
8. \(R_S \leftarrow \text{Decrypt}_{RSA}(R_S, PR_U)\)
9. If \(R = R_S\) then **Client Authenticate**
   \(CM_{sg} \leftarrow \text{Encrypt}_{AES}(\text{“Hello”, R})\)
   Send \(CM_{sg}\) to Server
10. Else RESET

**Server Side Continue:**
11. \(SM_{sg} \leftarrow \text{Decrypt}_{AES}(CM_{sg}, R)\)
12. If \(SM_{sg} = \text{“Hello”}\) then **Server Authenticate**
13. Else RESET
After receiving the ERN from server, client decrypts it with its private key and compares the received ERN with its generated ERN. If the both numbers are same then client authenticate the server and send “Hello” message to server encrypted with ERN using AES. On the server side, if the server correctly decrypts the “Hello” message using ERN then the server authenticates the client. After that step, communication session is started and ERN is used as shared symmetric key for that session. After the completion of authentication process, data management transactions are started for the current session. All transactions are encrypted with ERN by using AES. A special mechanism is followed for data transmission:

1. Each data chunk will be encrypted with $PR_U$ by using RSA, when it will be sent from client to server but the hash value of that data chunk will be encrypted with ERN by using AES. Hash will be encrypted with ERN in order to avoid replay attack.

2. When the data arrive at the server end, server decrypts the data with $PU_U$ and hash with ERN. After verifying the data integrity, data is encrypted again with $PU_U$ and stored on the server.

3. Whenever a client request for data, already encrypted data is sent to client with the hash value which is calculated from the encrypted data chunk. The attached hash value is also encrypted with ERN.

Fig 1: SKANS Authentication

SKANS uses CBC (Cipher Block Chaining) as a mode of operation for all data transmission and ERN is used as IV (Initial Vector) in CBC. Once the connection is established user starts communication with the server. To secure the communication SKANS uses a Counter Threat Mechanism (CTM). Idea behind CTM is such that if either user or server receives a message which have no meaning to receiving node then it is strong evidence that someone is trying to intercept in the communication. CTM works as follow:

1. During the communication each receiving end always tries to decrypt the received data with shared symmetric key.

2. If the data doesn’t decrypt successfully then in second attempt, it will try to decrypt either with PR (on user side) or PU (on server side).
3. A threat (e.g. data modification) is confirmed if the both decryption attempt result unsuccessful.

After discovering threat, threat detecting side (client or server); perform following actions to avoid further damage
1. The current session would be terminated immediately
2. Server would store all unacknowledged transactions for a specific period which will be prompted to the client on next login.
3. Threat detecting party also alerts the other party by sending a last warning message of current session encrypted with PR (on client side) or PU (on server side).

6. EFFICIENCY COMPARISON

In this section the efficiency of SKANS with the threats mentioned in section 2 has been compared. First of all the access policies proposed in SKANS ensures maximum security against the threats related with remote data storage as neither technical staff nor any intruder has any privilege to access users data or to modify it without user’s consent. Second level of our framework provides security against the threats associated with transactions. In SKANS, RN will be used as symmetric key for each session. The brute force attacks are unfeasible on this framework as every session uses a RN as a shared key. And the probability of brute forcing the RN is too low such that it requires 2 Supercomputers to process 3 days to brute force the RN so it seems almost computationally impossible. But Pseudo RNG (PRNG) generated random numbers are vulnerable to Algebraic attack. In order to avoid the algebraic attack, encrypted RN is used instead of algorithm generated RN. Against MITM attacks CTM provides efficient security as it not only detects any malicious activity but also close the attacked session immediately to avoid further damage. If any MITM tries to modify the message, server or client eventually would terminate the session so there is probably no chance that any MITM would create any danger during secure communication.

One advantage of SKANS on existing scenarios is that it is not browser dependent thus by using SKANS users have not to worry about the vulnerabilities browsers offer usually and can communicate with the server reliably. The last threat mentioned is the authentication and for this purpose SKANS uses a very delicate and attack-proof mechanism to grant authentication and there is a very rare chance that this mechanism would be breakable.

7. CONCLUSION AND FUTURE WORK

This paper discusses vulnerabilities related to cloud data storage and explained few existing solution. But security is not a simple issue which can be solved by proposing a framework however a careful analysis shows that SKANS presents a secure way of data storage in cloud computing. Theoretically examination shows that it removes most known vulnerabilities which attackers exploit to break in the user’s confidential data.

Although SKANS offer a reliable mechanism for data storage yet it requires more efforts to enhance existing capabilities. Future version of SKANS will constitute on a secure and efficient SKANS File System (SFS) to secure metadata and decrease latency in data access. Moreover solution will be generalized so that it can also be applied to other domains where data is being stored and accessed remotely. Our future research will
emphasize on enhancing the capabilities of SKANS to some more domains presented by the CSA.

REFERENCES

13. *eyeOS Developer Manual 0.1.0*
BUILDING CONSCIOUS CYBERNETIC ENTITIES USING QuBIC MODEL AND FRAMEWORK BASED ON UNIFIED THEORY OF MIND

Wajahat M. Qazi1,2* and Khalil Ahmed1
1 School of Computer Science, National College of Business Administration and Economics, Lahore, Pakistan
2 Department of Computer Science, GC University, Lahore, Pakistan. wmqazi@yahoo.com

ABSTRACT

Recently the scientific understanding of human conscious mind with reference to its ability to perform quantum computation has played a catalyst role in the technological advancement towards the development of conscious machines. As the result of this paradigm shift, new scientific and technological quest has emerged that address the issues of the performing integrative engineering that could integrate all possible parameter correlated with consciousness. These issues were recently addressed by QuBIC model that provides an integrative approach to cater essential correlates of consciousness as compared to its predecessor models. This study reports modifications in QuBIC model based on unified theory of mind. It also reports the development software frameworks associated with QuBIC model.

KEYWORDS
Artificial General Intelligence; Cybernetics; Machine Consciousness and Cognition; Quantum Neurocomputing; QuBIC; Unified Theory of Mind for Machine Consciousness

1. INTRODUCTION

Since the paradigm shift from artificial intelligence to general intelligence, many researchers have proposed various models that could assist in the engineering of conscious machines. Kinouchi [1] investigates consciousness and self with the objective to design autonomous adaptive system animal brain simulation using invisible computing and functional hierarchy. CODAM a neural network based on functional consciousness model was proposed by Taylor [2]. Another model was proposed by Bars and Franklin that Combines Global Workspace Theory and Franklin’s IDA architecture to model brain’s conscious and unconscious functions [3]. Global Workspace Theory [4] is one of oldest model assume a competition between several unconscious process. This competition is for entry into an architectural element known as Global Workspace. The winning contents are broadcasted to the competing processing changing their state. This is treated as the moment of conscious and its sequence is known as stream of consciousness. Global workspace has no implementation of phenomenal consciousness [5]. On the other hand, Gamez used spike neural networks to implement synthetic phenomenal consciousness in CRONOS [6]. Similarly Ron Sun initiated CLARION [7] project with the objective to explore the structure of the human mind by studying the interaction between implicit and explicit cognition. Existing work has played significant
role in understanding human cognition and the development of applied artifacts, which has led to the development of novel scientific theories.

Work in MC is based on cognitive theories, which are grounded on top of classical physics/computation and/or neural correlates of consciousness. The review of consciousness suggests other correlates of consciousness that discusses metaphysical, physical, quantum mechanical and religious aspects that may play a significant role in synthesis of consciousness. These various aspects and their significance are review elsewhere in literature [6,8,9,10].

Recently an effort was made to propose a solution called “cybernetic-cognition for machine consciousness”. This solution was formulated by unifying correlate of consciousness (unified theory of mind) in order to formulate a synergy among them and later it was extended for machine consciousness as unified theory of mind for machine consciousness (UTMMC) [11]. This extended version of the theory was then to build a model (QuBIC model) and associated frameworks (QuBIC and OpenQCF) for the development of artificial mind beings (AMB) [11]. Previously the developmental progress of QuBIC model and framework was reported [10] without its rationalization with reference to unified theory of mind [11]. This study reports progress made in the development and extension of QuBIC model and framework with reference to unified theory of mind.

2. UNIFIED THEORY OF MIND FOR MACHINE CONSCIOUSNESS


UTM stresses that the challenge in the designing of cognitive machine is to develop conceptual synergy among various cognitive knowledge domain (CKD). The aforementioned term CKD in this theory represents a collective reference for those knowledge bodies that implicitly or explicitly deals with the study of those elements of nature that could possibly play a significant role in the emergence of cognition (brain, mind, intelligence and consciousness). These knowledge domain are physics (classical and quantum), philosophy of mind, religion, mathematical formulations, biology, neuroscience, cognitive psychology, computational theories (classical and quantum), information theory (classical and quantum). It was formulated by reviewing following religious scientific and philosophical literatures: The Holy Quran, The Bible, Bhagavata Purana and Hua-Yen Buddahism, M-theory, Quantum Mechanics, My Big Toe and Theory of GOD. It addresses various issues related to framework of creation, causal relationship between multi-verse and human autonomy, relationship between GOD and creation with reference to cybernetics and unification of knowledge repository on scientific and religious grounds. Based on scientific and religious findings this theory is composed of proposes three components that unifies and develop synergy among the members of CKD. These components are dualistic-multiverse architecture (DMA) of creation, ‘GOD-Universe-Man and Cybernetics’ and unified knowledge repository, the detail of these components are reported elsewhere in literature.

Unified theory of mind for machine consciousness is extension of UTM. It extends GOD-Universe-Man cybernetic perspective of creation by incorporating machines in the equation thereby proposing machines as second order creation of GOD. Furthermore, it
also establish a relationship between relative knowledge of humans and artificial mind beings (machines with synthetic conscious mind developed using UTMMC and QuBIC model) to define relative knowledge of cybernetic organisms (cyborgs).

This theory concludes GOD created all creations in pairs resulting into duality. Moreover, GOD may use quantum interference to induce thoughts during awake or in dreams. It also deduces mind as information system and proposed the concept of seed based subjectively experienced knowledge (SSERK) based on Islamic philosophy of knowledge which explain the relationship between various forms of knowledge and their limitations.

3. QUANTUM AND BIO-INSPIRED INTELLIGENCE AND CONSCIOUSNESS

QuBIC is pragmatic component of aforementioned solution, cybernetic-cognition for machine consciousness that provides architectural abstraction and software framework for modeling and implementation of artificial mind beings. Since its previously published version[10] significant up-gradation were made in its components, figure 1 shows previous versions and figure 2 represents updated version which is according to UTMMC.

![QuBIC Model](image)

**Fig. 1: QuBIC Layered View**
3.1 QuBIC Model

The updated version of QuBIC model provides a dualistic perspective of cognition at all level of abstraction with reference to brain and mind. It provides mind-environment interface (MEI) unit as a metaphor of physical component and mind as a non-physical aspect of cognition called mind. The MEI unit is composed of actuators and sensors (internal and external) as shown in figure 2. The artificial mind component of QuBIC is composed of two cognitive processes, consciousness and unconsciousness. Like previous version [10] memory system, decision execution center, behavioral system and motivational system represent those cognitive systems, which perform functions with conscious and unconscious regimes. In previous version, meta-cognition was implemented in parallel to conscious and unconscious units. Whereas, in this updated version meta-cognition has dual representation in conscious and unconscious regime. Similarly, seed knowledge is now the part of unconscious regime according to Islamic philosophy of knowledge and SSERK described in UTMMC. Another significant change made QuBIC models is the addition of new cognitive processes, which are vital and play significant role in the synthesis of human-like cognition at conscious and unconscious levels. The components placed in unconscious layer are deed assessment module, circadian clock and circadian process monitor & control (CPMC). Furthermore, unconscious also encapsulates direct mind interface (DMI) and dreaming modules as shown in figure 2.

Circadian clock in QuBIC represents a synthetic version of circadian rhythm or body clock present in living beings, which is endogenously driven, approximately 24-hours cycle in biochemical, physiological or behavioral processes. In QuBIC model, circadian clock represents a discrete temporal implementation of various unconscious programmable processes specifically designed according to the objective function for a given artificial mind being. Whereas, CPMC is an internal monitoring and control unit for circadian clock. Its responsibility is to monitor the execution of circadian processes and to generate control signals accordingly. Deed assessment module in QuBIC model
Qazi and Ahmed

507

implements deed assessment vector space concept that utilizes seed knowledge to evaluate the performance in terms of deeds. If an action or a decision taken by an AMB violates the seed knowledge; the deed assessment module will evaluate the situation and compute the degree of violation as a feedback signal. This feedback signal is transmitted to emotion module in unconscious layer and to workspace in conscious block. The implementation of deed assessment module coupled with seed knowledge is a first step towards the implementation of human-like unconsciously triggered self-assessment and inner voice, which is, based on UTMMC.

3.2 QuBIC Framework

This framework is the implementation of QuBIC model as a software artifact, which exposes necessary API for development, analysis and evaluation of an AMB. It also acts as an execution engine, which symbolize and encapsulates the functionality of layer I of DMA, described in UTMMC.

As shown in figure 3, this framework is composed of three sub-systems. First sub-system represents the layer I of DMA from UTMMC, which provides various abstract classes, each of them representing cognitive process in QuBIC model as shown in figure 3. Remaining two sub-systems are OpenQCF and OpenNCF. OpenQCF is a collection of APIs that are used by QuBIC Layer I framework to implement basic quantum regularities used for communication between these cognitive components. OpenQCF was also used to develop quantum neural network and quantum genetic algorithms to implement various cognitive features. OpenNCF encapsulates classical and quantum neural network components that used as shown in figure 3.

4. TESTING AND EVALUATION OF CYBERNETIC-COGNITION FOR MACHINE CONSCIOUSNESS

The previous sections of this study reports UTMMC, QuBIC model and framework as a proposed solution for the development and analysis of artificial mind beings. This solution was test and evaluated by implementing JOHI, an artificial mind being capable to play paper-scissor-rock game based on its conscious and unconscious cognitive processes described earlier in the article. Quantitative and qualitative test were performed for the evaluation of JOHI to observe the presence of synthetic phenomenal
Building conscious cybernetic entities using QuBIC model…

consciousness. Quantitative analysis was performed using quantum mutual information [11] and qualitative analysis was performed using Metzingers constraints and Aleksander’s Axiom. The strategy used for qualitative analysis is similar to the methods used for the evaluation of CRONOS [6]. Both qualitative and quantitative tests verified the presence of phenomenal consciousness states in JOHI [11].

5. CONCLUSION

The development of software artifacts, which possess certain amount of consciousness and intelligence based on phenomenal and physical processes present in natural beings is challenging from engineering perspective, as it requires synergy among various cognitive knowledge domains. This study reports a solution for cybernetic-cognition for machine consciousness based on unification among cognitive knowledge domains. This unification was used to derive QuBIC model, which was later used to implement software framework. This solution was tested by implementing an artificial mind being, which was further analyzed for the existence of synthetic phenomenal consciousness using quantitative and qualitative analysis. The review analyses proposes that artificial mind beings developed using the solution reported here have the potential to possess certain amount of synthetic phenomenal conscious states.

REFERENCES

DETERMINANTS OF FOOD INFLATION IN PAKISTAN & EFFECTS OF SEASONAL ADJUSTMENT ON FORECASTING FOOD INFLATION

Sadia Batool1 and Javaid Shabbir2

1 State Bank of Pakistan, Karachi, Pakistan. Email: sadia.batool@sbp.org.pk
2 Department of Statistics, Quaid-e-Azam University, Islamabad, Pakistan. Email: js_qau@yahoo.com

ABSTRACT

Prices of food are sky-rocketing all over the world but the plight of the developing countries like Pakistan is quite miserable and the people who are already living below poverty line can hardly manage to make both ends meet. So in order to know the main causes of this fiasco we decided to conduct this research. The main focus of this project was to find out the factors influencing food inflation in Pakistan. After highlighting these variables a hybrid monetarist-structuralist model for food inflation was formulated. After performing standard unit root testing procedures a test was conducted for the presence of seasonal unit roots and all the variables needed seasonal adjustment. A detailed analysis of the long-run relationships between food prices and its determinants and causal nexus between them has been conducted both prior to and after seasonal adjustment using Demetra. A comparison of both the approaches shows that the results become more pragmatic when seasonal factor has been accounted for. Granger causality tests on the actual data show that food prices granger cause money supply, energy prices and wheat support price. Whereas exchange rate and value added in agriculture granger cause food prices. But when the data was adjusted for seasonality we came to know that two-way causality exists between value added in agriculture and food prices. Similarly Wheat Support Prices and Food Prices also have reverse causal links and both play equal part in determining each other. Besides two-way causal nexus we also notice that food prices granger cause energy prices and food exports. In addition to this it was found that Exchange rate, Fertilizer prices, Food imports and Money supply all granger cause food prices. Johansen cointegration test showed the existence of six cointegration vectors and this number was reduced to four when the data was adjusted for seasonality. Later on both the estimated models have been used for forecasting purposes in that case the model estimated with seasonally adjusted data performed much better than the other one.

1. INTRODUCTION

The dogs of hunger are not dead: some are sleeping, others are biting. In mid-2011 a food emergency unfolded in the Horn of Africa. Suddenly pictures of emaciated children were back in the media. Millions of people in East Africa are facing a food crisis caused by a perfect storm of severe drought, food price spikes, and conflict, and exacerbated by the vulnerability of people and communities across the region. Recent events in the Horn of Africa are a terrible reminder of the vulnerability of millions of poor around the world to weather and other shocks that interrupt their access to food. This humanitarian tragedy
Determinants of food inflation in Pakistan & effects…

highlights the need for information and the need for action. Addressing the problem of hunger requires information about where and why hunger is occurring. Information will not fill people’s stomachs, but policymakers and national and international agencies need it in order to take steps to ensure that people have access to sufficient and nutritious food. The broader task, though, is to take action to address the root causes of hunger and to reduce poor people’s vulnerability to shocks such as drought and food price spikes in the short, medium, and long term.

As far as the case of Pakistan is concerned the inflation rate in Pakistan was 11 percent in October of 2011. From 2003 until 2010, the average inflation rate in Pakistan was 10.15 percent reaching an historical high of 25.33 percent in August of 2008 and a record low of 1.41 percent in July of 2003. Rising inflation, especially of food, which is in double-digit for the last several years, is causing severe food insecurity in Pakistan, as millions of Pakistanis have limited access to food due to its high prices and the United Nations have time and again pinpointed this threat to the county. Although plenty of food is available in the market, yet its spiraling prices keep millions of people at bay and resultantly causing malnutrition among them According to the United Nations World Food Programme (WFP), more than 48 percent Pakistanis facing food insecurity. There is food available, but its affordability is a big issue. Interestingly, the government is not ready to accept that there is food insecurity in Pakistan, as it considers it as the “availability and not affordability”. The increased gap between the availability of food and its affordability has increased the incidence of food insecurity.

The central bank has kept its discount rate at 14 per cent in the past few months without any upward increase in it. As a result, this has led to partial increase in the non-food and non-energy core inflation, which hiked to double-digit of 10.4 per cent in June from 10.2 per cent in May. The house index rent rose in fiscal year 2011 by 7.29 per cent, medical care cost by 15.06 per cent and transportation fare by 14.38 per cent over the last year. The rise in transportation fare was driven mainly by the increase of diesel, petrol and CNG filling charges. Electricity charges also went up by more than 19 per cent in the whole year putting extra pressure on the pocket of consumers. Analysts point out two major factors aside from food products, which contributed to rise in the inflation in the year under review. The continued government borrowing from the central bank, which witnessed more than 74 per cent increase and totaled at Rs.685.3 billion during the year 2011. The analysts linked the hike in inflation with continued pressure of food prices, fiscal pressure due to security situation, adjustment in utility prices, continued depreciation of the rupee, which pushed upward cost of imported raw materials, goods and services, high mark-up rate and loss in productivity due to power shortages. The impact of imported inflation was also very high during the outgoing fiscal year, especially due to import of essential food items like sugar, edible oil, pulses and tea. The rise in the price of these products in international market and massive depreciation of rupee further led to increase in the price of these commodities in the domestic market. Aside from these, the supply shocks adversely impacted food inflation and were more visible in prices of heavy weights like pulses, rice, meat, milk, sugar and vegetables during the year under review. The surpassing of revenue target is believed to have helped the government to curtail the budget deficit to 5.2 pc of gross domestic products in the year ending June 30.
The rise in food insecurity in Pakistan has led to malnutrition. Interestingly, it is considered state of emergency if in a country or an area malnutrition is 15 percent or above. In Sindh, the malnutrition is at 22 percent which is a cause of concern, the sources said. The conflict and terrorism-ridden areas, especially Balochistan, Waziristan and far-fledged areas of the country are worse hit by the food insecurity and malnutrition. Pakistan’s agriculture growth is 1.2 percent in 2010/11, which is less than the official annual growth, while the rate of population of 2.1 percent; a sure sign of decline in per capita availability of food. Agriculturists said that Pakistan’s per capita grain area has shrunk between 38-56 percent from 1950 to 1998 and by 2050 this fall will go up to 55-63 percent. The projected growth for Pakistan to 350 million by 2050 will reduce its grain land per person from 0.08 hectares, at present, to 0.03 hectares, roughly the strip between 10-yard markers on a football field. Annual inflation went up by 13.92 percent in the fiscal 2010/11, leapt on the back of double-digit food inflation, rising oil prices in world market.

This study tries to find out the main causes of constantly rising food prices in Pakistan.

2. LITERATURE REVIEW

There are different schools of thought as to what causes inflation. The two most prevalent theories are the neo-classical/monetarists theory that inflation is driven by increases in the money supply, often used to finance government spending and the neo-Keynesian/structuralist view that inflation is the result of diminishing returns of productivity. The structuralist models of inflation emerged in the 1950. These were supported by Streeten (1962), Olivera (1964), Baumol (1967) and Maynard and Rijckeghem (1976). According to these models, supply-side factors like food prices, administered prices, wages and import prices are the determinants of inflation. The most widespread monetarist model of inflation is based on the quantity theory of money and asserts inflation is an increase in the supply of money at a rate greater than the expansion in the size of the economy. This is practically measured by comparing the GDP deflator to the rate of increase of the money supply, and setting the interest rate through the central bank to maintain a constant quantity of money.

Monetarists believe the most significant factor influencing inflation or deflation is how fast the money supply grows or shrinks. They consider fiscal policy, or government spending and taxation, as ineffective in controlling inflation. According to the famous monetarist economist Milton Friedman, "Inflation is always and everywhere a monetary phenomenon." Some monetarists, however, will qualify this by making an exception for very short-term circumstances.

Keynesian economic theory proposes that changes in money supply do not directly affect prices, and that visible inflation is the result of pressures in the economy expressing themselves in prices.

According to Neo-Keynesian economic theory there are three major types of inflation, as part of what Robert J. Gordon calls the "triangle model":

- Demands pull inflation - inflation due to high demand for GDP and low unemployment, also known as Phillips Curve inflation.
- Cost push inflation - nowadays termed "supply shock inflation", due to an event such as a sudden increase in the price of oil.
- Built-in inflation - induced by adaptive expectations, often linked to the "price/wage spiral" because it involves workers trying to keep their wages up with prices and then employers passing higher costs on to consumers as higher prices as part of a "vicious circle". Built-in inflation reflects events in the past, and so might be seen as hangover inflation. It is also known as "inertial" inflation, "inflationary momentum", and even "structural inflation".

These three types of inflation can be added up at any time to get an explanation of the current inflation rate. However, over time, the first two (and the actual inflation rate) affect the amount of built-in inflation: persistently high (or low) actual inflation leads to higher (lower) built-in inflation.

Within the context of the triangle model, there are two main elements: movements along the Phillips Curve, for example as unemployment rates fall, encouraging greater inflation, and shifts of that curve, as when inflation rises or falls at a given unemployment rate.

A major demand-pull theory centers on the supply of money: inflation may be caused by an increase in the quantity of money in circulation relative to the ability of the economy to supply (its potential output). This has been seen most graphically when governments have financed spending in a crisis by increasing the amount of currency in circulation to avoid the results of economic collapse, sometimes during wartime conditions. This has lead to hyperinflation where prices rise at extremely high rates in short periods of time in extreme cases.

A fundamental concept in such Keynesian analysis is the relationship between inflation and unemployment, called the Phillips curve. In classical Keynesian economics this model suggested that price stability was a trade off against employment. Therefore some level of inflation could be considered desirable in order to minimize unemployment. The Philips curve model described the US experience well in the 1960s, but failed to describe the combination of rising inflation and economic stagnation (sometimes referred to as stagflation) experienced in the 1970s. The modern use of the Phillips curve relates payroll growth to the general inflation rate, rather than relating the unemployment rate to the inflation rate, and suggests that tradeoffs between inflation and employment are based on the change in the rate of inflation, rather than the inflation rate itself.

In this model, increases in aggregate demand drive prices upwards, as suppliers are aware that they have pricing power, which leads to more people working, which leads to increased aggregate demand.

Thus, modern macroeconomics describes inflation using a Phillips curve that shifts (so the trade-off between inflation and unemployment changes) due to such matters as supply shocks and inflation becoming built into the normal workings of the economy. The former refers to such events as the oil shocks of the 1970s, while the latter refers to the price/wage spiral and inflationary expectations implying that the economy "normally"
suffers from inflation. Thus, the Phillips curve represents only the demand-pull component of the triangle model.

Another Keynesian concept is the potential output (sometimes called the "natural gross domestic product"), a level of GDP where the economy is at its optimal level of production, given institutional and natural constraints. This level of output corresponds to the NAIRU or the "natural" rate of unemployment or the full-employment unemployment rate. In this framework, the built-in inflation rate is determined endogenously (by the normal workings of the economy).

Austrian economics defines inflation as "an inflation of the supply of money". All other things being equal, inflation can be expected to cause an increase in prices, but exactly which prices are affected when and how much will depend on how the newly created money was introduced to the system, and how the new money spreads throughout the economy. Therefore, in the Austrian view, it is possible that changes in productivity (or other factors) will drive down the price of any arbitrarily chosen basket of goods, even in the presence of inflation (of the money supply). Thus fighting inflation is very simple in Austrian framework; just stop creating new money.... increasing prices can have many causes, and inflation (of the money supply) is but one. Austrian theory would dispute the idea that mild "deflation" (in the non-Austrian sense, i.e. a decline in the general price level) is something to be avoided. Instead, this is seen as a natural way of passing on productivity gains to those who have deferred consumption. Sudden contractions in the money supply can undoubtedly disrupt business plans which depended on continuous injections of new money, and bank runs are certainly unhealthy, but that doesn't mean all decreases in prices are inherently bad, or that continuing to inflate just to maintain inflation-dependent enterprises is the correct policy.

Supply-side economics asserts that inflation is always caused by either an increase in the supply of base money or a decrease in the demand for base money (or both). The value of money is seen as being purely subject to these two factors. Thus the inflation experienced during the Black Plague in medieval Europe is seen as being caused by a decrease in the demand for money as the volume of production and trade fell, while the inflation of the 1970s is regarded as been initially caused by an increased supply of money that occurred following the US exit from the Bretton Woods gold standard. Supply-side economics asserts that the money supply can grow without causing inflation as long as the demand for money also grows at the same rate.

Welfare economics takes the concept that the real purchasing power of an individual is measured in the basket of commodities that they can command. Therefore, it measures standard of living differently from GDP and price level, and instead uses the concept of "welfare" or happiness grounded in other measures. Neoclassical economics defines utility as being related to price, and therefore does not need to look at separate components of general welfare individually, only their aggregate price. This view is used by Marxian economists to argue that production and not consumption should be central to the definition of inflation.

Baek and Koo (2009) in their paper have used descriptive methods to investigate factors that affect the surge in U.S. food prices. In their paper they have made an attempt to examine the short- and long-run effects of changes in market factors such as prices of
energy and agricultural commodities and exchange rate on changes in U.S. food prices. For this purpose they employed the ARDL approach to co integration to estimate monthly price data from 1989 to 2008. The results show that agricultural commodity prices play a key role in affecting the short- and long-run behavior of U.S. food prices. They also find that energy prices and exchange rate have been significant factors influencing U.S. food prices in both the short- and long-run. This further suggests that the linkage between energy and agricultural commodity/food markets has been recently quite strong, due mainly to crop-based biofuel production.

Khan, Bukhari and Ahmed (2007) are of the view that the expansionary economic policies of the government and of the central bank (State Bank of Pakistan - SBP), which on one side resulted in impressive economic performance, stimulated a rise in the Consumer Price Index (CPI) on the other. This initiated a debate on the determinants of the recent inflation. Some blamed fiscal policy or monetary policy, while others blamed imported inflation, administered prices or mismanagement and lose control of the government. Their study, adopting an econometric framework, focused on the identification of the main determinants of inflation trends. Their findings show that the most important determinants of inflation in 2005-06 were adaptive expectations, private sector credit and rising import prices. Whereas, the fiscal policy’s contribution to inflation was minimal.

Gottschalk, Kalonji and Miyajima (2008) examined the determinants of inflation in Sierra Leone using a structural vector auto regression (VAR) approach to help forecast inflation for operational purposes. Despite data limitations, the paper accurately models inflation in Sierra Leone. As economic theory predicts, domestic inflation is found to increase with higher oil prices, higher money supply, and nominal exchange rate depreciation. The paper then employs a historical decomposition approach to pinpoint the sources of a marked decline in inflation in 2006 and assesses its forecasting properties.

3. METHODOLOGY AND DATA

3.1 Data and Model for Food Inflation

The following model has been formulated to capture the impacts of these variables on food inflation in Pakistan.

\[ Y_t = \alpha + \beta_1 M_{2t} + \beta_2 FI_t + \beta_3 FE_t + \beta_4 ER_t + \beta_5 EP_t + \beta_6 FP_t + \beta_7 VA_t + \beta_8 WSP_t + \xi_t \]  

(3.1)

where \( t = 1,2,3,\ldots,48 \) denotes time period from 2000-Q1 to 2011-Q4, \( \alpha \) is the constant intercept in the model and \( \beta \)'s are the slope coefficients for the respective variables. With respect to the signs of coefficients in equation (3.1), it is expected that \( \beta_i > 0 \), \( i=1,3,5,6,8 \) since an increase in money supply, energy prices, fertilizer prices, and wheat support prices results in an increase in food prices. As to the effect of exchange rate, the expected sign of \( \beta_4 \) is based on the definition of \( ER_t \). If \( ER_t \) is defined in a way that a decrease reflects a real depreciation of Rupee against major currencies, it is expected to be \( \beta_4 < 0 \); that is, the depreciation of Rupee leads to higher demand for agricultural commodity exports and thus prices. Similarly food imports and value added in agriculture are expected to have negative relationship with food prices in Pakistan. In (3.1) \( Y_t \)
represents Food price inflation and CPI food, compiled by the FBS has been used as a proxy for this variable.

\[ M_2_t \] is the money supply in time \( t \), data for money supply has been obtained from the State Bank of Pakistan.

\( F_I_t \) denotes the import prices of food and the relevant data has been acquired from the SBP.

\( F_E_t \) denotes the export prices of food and the relevant data has been acquired from the SBP.

\( E_R_t \) is real effective exchange rate at time \( t \). The data for \( E_R \) has also been acquired from the SBP, Statistics & DWH Department of the SBP compiles this data.

\( E_P_t \) stands for the energy prices and the respective data has been obtained from the data archive of the world bank.

\( F_P_t \) denotes fertilizer prices.

\( V_A_t \) shows the value added in agricultural production, the data employed has been acquired from the Agricultural Census Organization of Pakistan.

\( W_S_P_t \) denotes the Wheat Support Price announced by the government, the relevant data has been acquired from various publications of the Ministry of Finance.

### 3.2 Methodology

Applied economists and researchers have been aware of the problems they encountered in the analysis of the time series econometric models since mid 1980’s. These problems arise mainly due to the presence of unit roots in the data. Ignoring this fact and proceeding to estimate a regression model comprising of non-stationary variables at best ignores the important information about the underlying processes generating the data and at worst it is ought to lead to nonsensical (spurious) results. Owing to this reason it is incumbent upon the researcher to test for the presence of both zero-frequency and seasonal unit roots in the time series data and in case of their presence a judicious researcher is ought to choose an appropriate modelling startegy that best suits the nature of the data. De-trending is not appropriate and simply differencing the data to remove the non-stationary trend is only part of the answer. While the use of the differenced variables will avoid the problem of the spurious regression, it will also remove any long-run information that needs to be retained. But we also need to insure that this information reflects the co-movement of variables due to the underlying equilibrating tendencies of economic forces, rather than those due to common, but unrelated, time trends in data.

#### 3.2.1 Tests for Unit Roots in the Data

When considering long-run relationships, it becomes necessary to consider the underlying properties of the processes that generate time series variables. That is we must distinguish between stationary and non-stationary variables, since failure to do so can lead to a problem of spurious regression whereby the results suggest that there are statistically significant long-run relationships between the variables in the regression model when in fact all that is being obtained is evidence of contemporaneous correlations
rather than meaningful causal relations. Unless a non-stationary variable combines with other non-stationary series to form a stationary co-integration relationship, the regressions involving that variable can falsely imply the existence of a meaningful relationship.

In principle it is important to test the order of integration of each variable in a model, to establish whether it is non-stationary and how many times it needs to be differenced to result in a stationary series.

Testing for the unit root in a series is not a straightforward task, rather the following issues need to be resolved first:

a. We have to take this possibility into consideration that the unknown underlying data-generating process (dgp) may include a time trend (stochastic or deterministic).

b. The dgp may be more complicated than a simple autoregressive AR(1) process, and indeed may involve Moving Average terms.

c. It is known that when dealing with finite samples the standard tests for unit roots are biased toward accepting the null hypothesis of non-stationarity when the true dgp is in fact stationary (i.e. there is problem with the power of the test).

d. There is concern that an undetected structural break in the series may lead to under-rejecting the null.

e. Quarterly data needs to be tested for seasonal unit roots in addition to the usual test for unit roots at the zero frequency level.

3.2.2 Graphical Evidence to Gauge Non-Stationarity

The graphical inspection of data reveals a lot of information about the trends present in the data and by having the graphical picture of the data we can form the first judgement about it. As far as non-stationarity of the time series data employed is concerned prior to using the conventional unit roots testing procedures we can use many graphical techniques like correlograms etc. to get an idea about whether the corresponding time series is stationary or non-stationary.

3.2.3 Augmented Dickey-Fuller (ADF) Test

To test for the presence of unit roots in the time series employed the ADF test has been used along with other econometric tools. Let’s assume that a given time series say Yt follows a pth order AR process:

\[ y_t = \Psi_1 y_{t-1} + \Psi_2 y_{t-2} + \ldots + \Psi_p y_{t-p} + u_t \]

\[ \Delta y_t = \Psi^* y_{t-1} + \Psi_1 \Delta y_{t-1} + \Psi_2 \Delta y_{t-2} + \ldots + \Psi_{p-1} \Delta y_{t-p+1} + u_t \]

where \( \Psi^* = (\Psi_1 + \Psi_2 + \ldots + \Psi_p) - 1 \). If \( \Psi^* = 0 \), against the alternative \( \Psi^* < 0 \), then \( y_t \) contains a unit root. To test the null hypothesis the test statistic \( t = \Psi^*/S.E.(\Psi^*) \) is calculated.

The model illustrated in (3.2) can be extended to allow for the possibility that the dgp contains deterministic components. The model needed to test for the null hypothesis of a stochastic trend (non-stationary) against a deterministic trend (stationary) is as follows:
\[ \Delta y_t = \Psi^* y_{t-1} + \Psi_1 \Delta y_{t-1} + \Psi_2 \Delta y_{t-2} + \ldots + \Psi_{p-1} \Delta y_{t-p+1} + \mu + \gamma t + u_t \] (3.3)

The augmented model can be extended even further to take into account MA parts in the ut. It is generally believed that MA terms are present in many macroeconomic time series after first differencing (e.g., time average data, an index of stock prices with infrequent trading for a subset of the index, the presence of errors in the data, etc.)

### 3.2.4 Selection of Lag Length for ADF Test

Before proceeding towards unit root testing via model (3.3) an important issue that needs to be resolved is the selection of appropriate lag length \( p \). It is of very crucial importance to select appropriate lag length; too few lags may result in over-rejecting the null when it is true (i.e., adversely affecting the size of the test), while too many lags may reduce the power of the test (since unnecessary nuisance parameters reduce the effective number of observations available).

Banerjee et al. (1993) favour a generous parameterization since if too many lags are present the regression is free to set them to zero at the cost of some loss in efficiency, whereas too few lags implies some remaining autocorrelation and hence the inapplicability of even the asymptotic distributions.

Suggested solutions to the choice of lag length \( p \) in (3.3) involve using a model selection procedure that tests to see if an additional lag is significant (e.g., if it increases the value of \( \bar{R}^2 \)). However, it is shown in Harris (1992) that maximizing \( \bar{R}^2 \) to choose the value of \( p \) in the ADF test proved to be unsatisfactory; Monte Carlo experiments undertaken using various dgps (e.g., ARMA,MA and AR) suggested that there were problems with the size of this form of the ADF test. Rather choosing a fairly generous value of \( p \) using a formula suggested by Schwert (1989), that allows the order of autoregression to grow with the sample size \( T \), resulted in a test with size close to its nominal value (i.e., 10%, 5% and 1%). This is consistent with Banerjee et al. (1993), and thus Harris (1992) suggested that the lag length should normally be chosen on the basis of the formula reported in Schwert (1989) that is \( l_{12} = \text{int} \{ 12(T/100)^{1/4} \} \), where \( T \) represents the size of a time series and \( \text{int} \) shows that only the integer part of the calculation is taken as lag length.

Ng and Perron (1995) argue in favour of a general-to-specific sequential t-test for the significance of the last lag that has the ability to yield higher values of lag length, \( p \) rather than standard lag length selection criteria. Hence this reduces size distortions, but this approach also tends to overparameterize when MA errors are not important. Weber (2001) argues in favour of a specific –to-general approach whereby \( p \) is initially set at a low value, and then F-tests are conducted for eliminating longer lags from the model (i.e., the lag length in the ADF regression is set at the smallest \( p \) such that all longer lags up to \( p_{\text{max}} \), maximum possible lag length that is decided by Schwert (1989) formulae, are jointly significant.

Khim and Liew (2004) found that Akaike’s information criterion (AIC) and final prediction error (FPE) are superior than the other criteria under study in the case of small sample (60 observations and below), in the manners that they minimize the chance of under estimation while maximizing the chance of recovering the true lag length. One
immediate econometric implication of their study is that as most economic sample data can seldom be considered “large” in size, AIC and FPE are recommended for the estimation the autoregressive lag length.

Asghar and Abid (2009) have compared various lag length selection criteria (Akaike Information Criterion, Schwarz Information Criterion, Hannan-Quinn Criterion, Final Prediction Error, Corrected version of AIC) for three different cases i.e. under normal errors, under non-normal errors and under structural break by using Monte Carlo simulation. They concluded that SIC is the best for large samples and no criteria is useful for selecting true lag length in presence of regime shifts or shocks to the system. Hence as recommended by Khim and Liew (2004) we have used the AIC for selection of lag length since the sample size for the data employed was small.

3.2.5 Seasonal Unit Root Tests

Certain variables (e.g., consumption, spending) exhibit strong seasonal patterns that account for a major part of the total variation in the data and that are therefore important when model-building.

A stationary seasonal process is conventionally modelled using seasonal dummies, that allow some variation, but no persistent change in the seasonal pattern over time, the drifting of the quarters over time may indicate that deterministic seasonal modelling is inadequate. That is the seasonal processes may be non-stationary if there is a varying and changing seasonal pattern over time. Such processes can not be captured using deterministic seasonal dummies since the seasonal component drifts substantially over time; instead such a series needs to be seasonal-differenced to achieve stationarity. This is complicated than considering the possibility of a unit root (non-stationarity) at the zero frequency since there are four different unit roots possible in a seasonal process. To see this consider seasonal differencing quarterly data using the seasonal difference operator

$$\Delta_4 y_t = (1-L^4)y_t = y_t - y_{t-4} .$$

where \((1-L^4)\) can be factorized as:

$$(1-L^4) = (1-L)(1+L+L^2 + L^3) = (1-L)(1+L^2)(1+L^2) = (1-L)(1+L)(1-iL)(1+iL)$$

where \(i = \sqrt{-1}\) is iota and \(L\) is the lag operator whereas \(\Delta\) is the difference operator. With each unit root corresponding to a different cycle in the time domain. In equation (3.4) \((1-L)\) is the standard unit root at the zero frequency. The remaining unit roots are obtained from the Moving Averages seasonal filter \(S(L) = (1+L+L^2+L^3)\) and these correspond to the two-quarter (half-yearly) frequency \((1+L)\) and a pair of complex conjugate roots at the four-quarter (annual) frequency \((1+iL)\).

To incorporate seasonnal integration into the definition of integration at the zero frequency, it is useful to note as above that seasonal differencing involves using \((1-L)\) to difference at the zero frequency \(d\) in order to remove the zero frequency unit roots, and using the seasonal filter \(S(L)\) to difference at the seasonal frequency \(D\), in order to remove the seasonal unit roots. Thus, it is said that the stochastic process \(y_t\) is integrated of order
d and D, I(d,D), if the series is stationary after first period differencing d times and seasonal-differencing D times. To test the number of seasonal unit roots in a univariate time series, the common approach is to use the procedure described in Hylleberg, Engle, Granger and Yoo (1990), abbreviated as HEGY (1990). The following regression is estimated using OLS with tests of the amounting to tests for various unit roots that may be present in the series:

\[
\Delta_d y_t = \beta_1 D_{1t} + \beta_2 D_{2t} + \beta_3 D_{3t} + \beta_4 D_{4t} + \pi_1 Z_1 y_{t-1} + \pi_2 Z_2 y_{t-1} + \pi_3 Z_3 y_{t-2} \\
+ \pi_4 Z_3 y_{t-1} + \sum_{i=1}^{p-1} \alpha_i \Delta_d y_{t-1} + \mu t + \epsilon_t
\]

(3.5)

where \(D_{qt}\) is the zero/one dummy corresponding to quarter q and where:

\[
Z_1 = (1 + L + L^2 + L^3) \\
Z_2 = -(1 - L + L^2 - L^3) \\
Z_3 = -(1 - L^2)
\]

If \(\pi_1 = 0\) then \(1-L\) filter is required to remove unit root present at zero frequency; if \(\pi_2 = 0\) then \(1+L\) filter is needed since the series contains unit root at half-yearly frequency; and when \(\pi_3 = \pi_4 = 0\) the series contains unit root at four-quarter(annual) frequency and \(1+L^2\) filter is needed to remove unit root. In case if \(\pi_1 = \pi_2 = \pi_3 = \pi_4 = 0\) then all the seasonal and non-seasonal all the unit roots are present in the given series.

3.2.6 Seasonal Adjustment of the Data

To adjust the time series data for seasonal variation a software package known as Demetra has been used.

Demetra was designed by Eurostat to provide ease of access for non-specialists to TRAMO, SEATS and X-12-ARIMA (specialized techniques for seasonal adjustment of the time series data) and it improves largely their user-friendliness. Demetra is a tool for statistical production in a large-scale environment imposing a recognized seasonal adjustment policy. It seasonally adjusts large-scale sets of time series, checks the quality of the results, improves the stability of the models, automatically improves rejected adjustments and assists the users in all the treatment. Additionally, it allows detailed analysis on single time series. The interface mainly uses the statistical algorithms included in the SA methods X-12-ARIMA and TRAMO/SEATS. It is a fully menu driven package, using general statistical vocabulary for parameters, models and functions except for very advanced usage.

3.2.7 Test for Granger Causality

While conducting empirical analysis based on the variables selected it was presumed that the causal relationships can be present between them. So to get an idea about the causal nexus between the variables the test for Granger causality has been conducted using E views package for time series and econometrics.
Assume that the information set $F_t$ has the form, where $x_t$ and $z_t$ are vectors (that includes scalars of course) and $z_t$ usually will include $y_t$ and $z_t$ may or may not include other variables than $y_t$. We say that $x_t$ is Granger causal for $y_t$ with respect to $F_t$ if the variance of the optimal linear predictor of $y_{t+h}$ based on $F_t$ has smaller variance than the optimal linear predictor of $y_{t+h}$ based on $z_t; z_{t+1}, ...$ for any $h$. In other words $x_t$ is Granger causal for $y_t$ if $x_t$ helps predict $y_t$ at some stage in the future. Often we have that $x_t$ Granger causes $y_t$ and $y_t$ Granger causes $x_t$. Granger causality measures whether one thing happens before another thing and helps predict it - and nothing else.

Assume that our data can be described by the model

$$
\begin{bmatrix}
  y_t \\
  z_t \\
  x_t
\end{bmatrix}
= 
\begin{bmatrix}
  \mu_1 \\
  \mu_2 \\
  \mu_3
\end{bmatrix}
+ 
\begin{bmatrix}
  A_{11} & A_{12} & A_{13} \\
  A_{21} & A_{22} & A_{23} \\
  A_{31} & A_{32} & A_{33}
\end{bmatrix}
\begin{bmatrix}
  y_{t-1} \\
  z_{t-1} \\
  x_{t-1}
\end{bmatrix}
+ ...
+ 
\begin{bmatrix}
  A_{11}^k & A_{12}^k & A_{13}^k \\
  A_{21}^k & A_{22}^k & A_{23}^k \\
  A_{31}^k & A_{32}^k & A_{33}^k
\end{bmatrix}
\begin{bmatrix}
  y_{t-k} \\
  z_{t-k} \\
  x_{t-k}
\end{bmatrix}
+ 
\begin{bmatrix}
  u_{11} \\
  u_{21} \\
  u_{31}
\end{bmatrix}
$$

Also assume that

$$
\Sigma_u = 
\begin{bmatrix}
  \Sigma_{11} & \Sigma_{12} & \Sigma_{13} \\
  \Sigma_{12} & \Sigma_{22} & \Sigma_{23} \\
  \Sigma_{13} & \Sigma_{23} & \Sigma_{33}
\end{bmatrix}
$$

(3.8)

This model is a totally general VAR-model - only the data vector has been partitioned in 3 subsectors - the $y_t$ and the $x_t$ vectors between which we will test for causality and the $z_t$ vector (which may be empty) which we condition on. In this model it is clear that $x_t$ does not Granger cause $y_t$ with respect to the information set generated by $z_t$ if either $A_{13}^i = 0$ and $A_{23}^i = 0$; $i = 1, ..., k$ or $A_{13}^i = 0$ and $A_{12}^i = 0$; $i = 1, ..., k$.

This is the way we will test for Granger causality. Usually we use the VAR approach if we have an econometric hypothesis of interest that states that $x_t$ Granger causes $y_t$ but $y_t$ does not Granger cause $x_t$. Sims (1972) is a paper that became very famous because it showed that money Granger causes output, but output does not Granger cause money. Later Sims showed that this conclusion did not hold if interest rates were included in the system. This also shows the major drawback of the Granger causality test namely the dependence on the right choice of the conditioning set. In reality one can never be sure that the conditioning set has been chosen large enough (and in short macro-economic series one is forced to choose a low dimension for the VAR model), but the test is still a useful (although not perfect) test.

### 3.2.8 Cointegration among the variables

The Johansen technique is an essential tool for applied economists who wish to estimate time series models. The implication that non-stationary variables can lead to spurious regression unless at least one cointegration vector is present means that some form of testing for cointegration is almost mandatory. Engle-Granger approach can be used for testing the cointegration between the variables in a time series model. But the problem is that it is the single equation approach and tests for the presence of only one
vector at a time. But if there are \( n > 2 \) variables in the model then there can be more than one cointegration vector. It is possible for \( n-1 \) linearly independent cointegration vectors to exist, and only when \( n = 2 \) it is possible to show that cointegration vector is unique.

Assuming that there is only one cointegration vector, when in fact there are more, leads to inefficiency in the sense that we can only obtain a linear combination of these vectors when estimating a single equation model. However, the drawbacks of this approach extend beyond its inability to validly estimate the long-run relationship between variables. Even if there is only one cointegration relationship, estimating a single equation is potentially inefficient. Information is lost unless each endogenous variable appears on the left-hand side of the estimated equations in the multivariate model, except in the case where all the right-hand-side variables in the cointegration vector are weakly exogenous.

Considering the problems that can be encountered when using Engle Granger approach it was decided to use the Johansen approach to test for the presence of multiple cointegration relationships between the variables. The issues facing the one who wishes to implement the technique include, inter alia:

a. Testing the order of integration of each variable that enters the multivariate model;
b. Setting the appropriate lag-length of the vector autoregression (VAR) model and determining whether the system should be conditioned on any predetermined \( I(0) \) variables including dummies to take account of possible policy interventions;
c. Testing for reduced rank, including the issue of testing whether the system should be treated as \( I(2) \) rather than an \( I(1) \) system;
d. Identifying whether there are trends in the data and whether deterministic variables (an intercept or trend) should enter the cointegration space or not;
e. Testing for weak exogeneity;
f. Testing for the unique cointegration vectors and joint tests involving restrictions on \( \alpha \) and \( \beta \).

**Johansen Cointegration Procedure**

The general form of Vector Auto regression (VAR) involving up to \( k \) lags of \( z_t \) can be written as following:

\[
z_t = A_1 z_{t-1} + \ldots + A_k z_{t-k} + u_t \quad u_t \sim IN(0, \Sigma)
\]

where \( z_t \) is \((n \times 1)\) and each of \( A_i \) is \((n \times n)\) matrix of parameters. This type of VAR model has been advocated most notably by Sims (1980) as a way to estimate dynamic relationships among jointly endogenous variables without imposing a priori restrictions. Equation 3.6 can be reformulated into a VECM form:

\[
\Delta z_t = \Gamma_1 \Delta z_{t-1} + \ldots + \Gamma_{k-1} \Delta z_{t-k+1} + \Pi z_{t-k} + u_t
\]

where \( \Gamma_i = -(I - A_i - \ldots - A_k)(i = 1, \ldots, k-1) \) and \( \Pi = -(I - A_1 - \ldots - A_k) \). This way of specifying the system contains information on both the short-run and long-run adjustments to changes in \( z_t \) via the estimates of \( \Gamma \) and \( \Pi \), respectively. \( \Pi = \alpha \beta' \), where \( \alpha \) represents the speed of adjustment to disequilibrium and \( \beta \) is a matrix of long-run
coefficients such that the term $\beta' z_{t-k}$ embedded in (3.7) represents up to (n-1) co
integration relationships in the multivariate model.

Assuming $z_t$ is a vector of non-stationary I(1) variables, then all the terms in 3.7 that
involve $\Delta z_{t-i}$ are I(0) while $\Pi z_{t-k}$ must also be stationary for stochastic error term in 3.7
to be white noise. There are three instances when the requirement that $\Pi z_{t-k} \sim I(0)$ is
met; when all the variables in $z_t$ are stationary, which is an uninteresting case in the
present context since it implies that there is no problem of spurious regression and
appropriate modeling strategy is to estimate the standard Sims-type VAR in levels (i.e.,
equation (3.6). The second instance is when there is no co integration at all, implying that
there are no linear combinations of the $z_t$ that are I(0), and consequently $\Pi$ is an $(n \times n)$
matrix of zeros. In this case, the appropriate model is a VAR in first-difference involving
no long-run elements. The third way for $\Pi z_{t-k}$ to be I(0) is when there exists up to
(n-1) co integration relationships: $\beta' z_{t-k} \sim I(0)$. In this instance $r \leq (n-1)$ co
integration vectors exist in $\beta$ (i.e., r columns of $\beta$ form r linearly independent
combinations of the variables in $z_t$, each of which is stationary), together with (n-r)
non-stationary vectors (i.e., r columns of $\beta$ form I(1) common trends). Only the co
integration vectors in $\beta$ enter 3.7, otherwise $\Pi z_{t-k}$ would not be I(0), which implies that
the last (n-r) columns of $\alpha$ are insignificantly small (i.e., effectively zero). Consequently,
testing for co integration amounts to a consideration of the rank of $\Pi$ (i.e., finding the
number of r linearly independent columns in $\Pi$). To summarize, if $\Pi$ has full rank (i.e.,
there are r = n linearly independent columns), then the variables in $z_t$ are I(0), while if
the rank of $\Pi$ is zero, then there are no co integration relationships. Neither of these cases
is particularly interesting. More usually $\Pi$ has reduced rank (i.e., there are $r \leq (n-1)$
cointegration vectors present). Tests for reduced rank can be conducted using trace
statistic $\lambda_{\text{trace}}$ and maximal eigen value or $\lambda_{\text{max}}$ statistic.

4. EMPIRICAL EVIDENCE

4.1 Tests for Unit Roots

As mentioned earlier in the chapter 4, Augmented Dickey-Fuller test has been used to
determine whether the variables are stationary or they need some differencing to become
stationary. Details of the ADF test have been highlighted in the previous chapter.

The results obtained by the application of ADF test are presented in the following
table:
Table 4.1: ADF Test at Level

<table>
<thead>
<tr>
<th>Variable</th>
<th>Intercept &amp; Trend both Not Included</th>
<th>Intercept Included</th>
<th>Intercept &amp; Trend both Included</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>t-statistic</td>
<td>Prob.*</td>
<td>t-statistic</td>
</tr>
<tr>
<td>$Y_t$</td>
<td>3.346404</td>
<td>0.9997</td>
<td>1.917545</td>
</tr>
<tr>
<td></td>
<td>1% level</td>
<td>-2.6162</td>
<td>-3.58115</td>
</tr>
<tr>
<td></td>
<td>5% level</td>
<td>-1.94814</td>
<td>-2.92662</td>
</tr>
<tr>
<td></td>
<td>10% level</td>
<td>-1.61232</td>
<td>-2.60142</td>
</tr>
<tr>
<td>$M_2_t$</td>
<td>13.5754</td>
<td>1</td>
<td>4.400542</td>
</tr>
<tr>
<td></td>
<td>1% level</td>
<td>-2.61509</td>
<td>-3.57772</td>
</tr>
<tr>
<td></td>
<td>5% level</td>
<td>-1.94798</td>
<td>-2.92517</td>
</tr>
<tr>
<td></td>
<td>10% level</td>
<td>-1.61241</td>
<td>-2.60066</td>
</tr>
<tr>
<td>$E_{R_t}$</td>
<td>0.127277</td>
<td>0.718</td>
<td>-1.72833</td>
</tr>
<tr>
<td></td>
<td>1% level</td>
<td>-2.61509</td>
<td>-3.57772</td>
</tr>
<tr>
<td></td>
<td>5% level</td>
<td>-1.94798</td>
<td>-2.92517</td>
</tr>
<tr>
<td></td>
<td>10% level</td>
<td>-1.61241</td>
<td>-2.60066</td>
</tr>
<tr>
<td>$E_{P_t}$</td>
<td>0.60371</td>
<td>0.8431</td>
<td>-0.8993</td>
</tr>
<tr>
<td></td>
<td>1% level</td>
<td>-2.61736</td>
<td>-3.58474</td>
</tr>
<tr>
<td></td>
<td>5% level</td>
<td>-1.94831</td>
<td>-2.92814</td>
</tr>
<tr>
<td></td>
<td>10% level</td>
<td>-1.61223</td>
<td>-2.60223</td>
</tr>
<tr>
<td>$F_{P_t}$</td>
<td>1.618526</td>
<td>0.9724</td>
<td>0.199129</td>
</tr>
<tr>
<td></td>
<td>1% level</td>
<td>-2.61985</td>
<td>-3.59246</td>
</tr>
<tr>
<td></td>
<td>5% level</td>
<td>-1.94869</td>
<td>-2.9314</td>
</tr>
<tr>
<td></td>
<td>10% level</td>
<td>-1.61204</td>
<td>-2.60394</td>
</tr>
<tr>
<td>$F_{I_t}$</td>
<td>0.202402</td>
<td>0.7406</td>
<td>-1.27586</td>
</tr>
<tr>
<td></td>
<td>1% level</td>
<td>-2.61509</td>
<td>-3.57772</td>
</tr>
<tr>
<td></td>
<td>5% level</td>
<td>-1.94798</td>
<td>-2.92517</td>
</tr>
<tr>
<td></td>
<td>10% level</td>
<td>-1.61241</td>
<td>-2.60066</td>
</tr>
<tr>
<td>$F_{E_t}$</td>
<td>4.7668</td>
<td>1</td>
<td>3.056326</td>
</tr>
<tr>
<td></td>
<td>1% level</td>
<td>-2.62724</td>
<td>-3.61559</td>
</tr>
<tr>
<td></td>
<td>5% level</td>
<td>-1.94986</td>
<td>-2.94115</td>
</tr>
<tr>
<td></td>
<td>10% level</td>
<td>-1.61147</td>
<td>-2.60907</td>
</tr>
<tr>
<td>$W_{S_{P_t}}$</td>
<td>1.750455</td>
<td>0.9792</td>
<td>-0.44145</td>
</tr>
<tr>
<td></td>
<td>1% level</td>
<td>-2.61509</td>
<td>-3.57772</td>
</tr>
<tr>
<td></td>
<td>5% level</td>
<td>-1.94798</td>
<td>-2.92517</td>
</tr>
<tr>
<td></td>
<td>10% level</td>
<td>-1.61241</td>
<td>-2.60066</td>
</tr>
<tr>
<td>$V_{A_t}$</td>
<td>1.914759</td>
<td>0.9851</td>
<td>-0.0618</td>
</tr>
<tr>
<td></td>
<td>1% level</td>
<td>-2.62561</td>
<td>-3.61045</td>
</tr>
<tr>
<td></td>
<td>5% level</td>
<td>-1.94961</td>
<td>-2.93899</td>
</tr>
<tr>
<td></td>
<td>10% level</td>
<td>-1.61159</td>
<td>-2.60793</td>
</tr>
</tbody>
</table>

It is quite evident from the results obtained in table 4.1 that the ADF test when applied on the levels of all the variables indicates the presence of unit roots in the data. P-values for all the three cases are very high which is an indication of the highly significant null hypothesis of unit root.
Table 4.2: ADF Test at First Difference

<table>
<thead>
<tr>
<th>Variable</th>
<th>Intercept &amp; Trend both Not Included</th>
<th>Intercept Included</th>
<th>Intercept &amp; Trend both Included</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>t-statistic</td>
<td>Prob.</td>
<td>t-statistic</td>
</tr>
<tr>
<td>ΔY_t</td>
<td>-2.45405</td>
<td>0.0152</td>
<td>-4.78507</td>
</tr>
<tr>
<td></td>
<td>1% level</td>
<td>-2.6162</td>
<td>-4.17058</td>
</tr>
<tr>
<td></td>
<td>5% level</td>
<td>-1.94814</td>
<td>-3.51074</td>
</tr>
<tr>
<td></td>
<td>10% level</td>
<td>-1.61232</td>
<td>-3.18551</td>
</tr>
<tr>
<td>ΔM2_t</td>
<td>-0.8505</td>
<td>0.342</td>
<td>-2.35624</td>
</tr>
<tr>
<td></td>
<td>1% level</td>
<td>-2.61736</td>
<td>-3.58474</td>
</tr>
<tr>
<td></td>
<td>5% level</td>
<td>-1.94831</td>
<td>-2.92814</td>
</tr>
<tr>
<td></td>
<td>10% level</td>
<td>-1.61223</td>
<td>-2.60223</td>
</tr>
<tr>
<td>ΔER_t</td>
<td>-5.8771</td>
<td>0.000</td>
<td>-5.80871</td>
</tr>
<tr>
<td></td>
<td>1% level</td>
<td>-2.61736</td>
<td>-3.58474</td>
</tr>
<tr>
<td></td>
<td>5% level</td>
<td>-1.94831</td>
<td>-2.92814</td>
</tr>
<tr>
<td></td>
<td>10% level</td>
<td>-1.61223</td>
<td>-2.60223</td>
</tr>
<tr>
<td>ΔFP_t</td>
<td>-6.31668</td>
<td>0.000</td>
<td>-6.42567</td>
</tr>
<tr>
<td></td>
<td>1% level</td>
<td>-2.61736</td>
<td>-3.58474</td>
</tr>
<tr>
<td></td>
<td>5% level</td>
<td>-1.94831</td>
<td>-2.92814</td>
</tr>
<tr>
<td></td>
<td>10% level</td>
<td>-1.61223</td>
<td>-2.60223</td>
</tr>
<tr>
<td>ΔFI_t</td>
<td>-1.33628</td>
<td>0.1653</td>
<td>-2.19465</td>
</tr>
<tr>
<td></td>
<td>1% level</td>
<td>-2.61985</td>
<td>-3.59246</td>
</tr>
<tr>
<td></td>
<td>5% level</td>
<td>-1.94869</td>
<td>-2.9314</td>
</tr>
<tr>
<td></td>
<td>10% level</td>
<td>-1.61204</td>
<td>-2.60394</td>
</tr>
<tr>
<td>ΔFE_t</td>
<td>-7.86029</td>
<td>0.000</td>
<td>-7.9252</td>
</tr>
<tr>
<td></td>
<td>1% level</td>
<td>-2.6162</td>
<td>-3.58115</td>
</tr>
<tr>
<td></td>
<td>5% level</td>
<td>-1.94814</td>
<td>-2.92662</td>
</tr>
<tr>
<td></td>
<td>10% level</td>
<td>-1.61232</td>
<td>-2.60142</td>
</tr>
<tr>
<td>ΔWSP_t</td>
<td>-10.4629</td>
<td>0.000</td>
<td>-11.3977</td>
</tr>
<tr>
<td></td>
<td>1% level</td>
<td>-2.61736</td>
<td>-3.58474</td>
</tr>
<tr>
<td></td>
<td>5% level</td>
<td>-1.94831</td>
<td>-2.92814</td>
</tr>
<tr>
<td></td>
<td>10% level</td>
<td>-1.61223</td>
<td>-2.60223</td>
</tr>
<tr>
<td>ΔVA_t</td>
<td>-1.78117</td>
<td>0.0713</td>
<td>-7.45509</td>
</tr>
<tr>
<td></td>
<td>1% level</td>
<td>-2.61985</td>
<td>-3.58115</td>
</tr>
<tr>
<td></td>
<td>5% level</td>
<td>-1.94869</td>
<td>-2.92662</td>
</tr>
<tr>
<td></td>
<td>10% level</td>
<td>-1.61204</td>
<td>-2.60142</td>
</tr>
<tr>
<td></td>
<td>-5.51205</td>
<td>0.000</td>
<td>-6.01357</td>
</tr>
<tr>
<td></td>
<td>1% level</td>
<td>-2.62561</td>
<td>-3.61045</td>
</tr>
<tr>
<td></td>
<td>5% level</td>
<td>-1.94961</td>
<td>-2.93899</td>
</tr>
<tr>
<td></td>
<td>10% level</td>
<td>-1.61159</td>
<td>-2.60793</td>
</tr>
</tbody>
</table>

Table 4.2 gives the results of the ADF test after each variable has been differenced once and it is indicated by the results obtained by the ADF test that all the variables become stationary after they have been differenced. So all of the variables employed have the same order of integration i.e., all variables are I(1).
4.2 Seasonal Unit Root Tests

To take into account seasonality of the data while testing for unit roots HEGY procedure highlighted in detail in chapter 4 has been employed. As per HEGY procedure the following regression has been estimated for each variable separately:

\[ \Delta_y_t = \beta_1 D_1 + \beta_2 D_2 + \beta_3 D_3 + \beta_4 D_4 + \pi_1 Z_{1,y} + \pi_2 Z_{2,y} + \pi_3 Z_{3,y} + \pi_4 Z_{4,y} + \sum_{i=1}^{4} \alpha_i \Delta_y_{t-i} + \mu + \varepsilon, \]

If \( \pi_1 = 0 \) then only zero frequency unit root is present that can be removed by simply differencing the data; if \( \pi_2 = 0 \) then the series contains unit root at half-yearly frequency; and when \( \pi_3 = \pi_4 = 0 \) the series contains unit root at four-quarter(annual) frequency. In case if \( \pi_1 = \pi_2 = \pi_3 = \pi_4 = 0 \) then all the seasonal and non-seasonal unit roots are present in the given series.

The results of the seasonal unit root test are given in the following table:

<table>
<thead>
<tr>
<th>Variable</th>
<th>( H_0: \pi_1 = 0 )</th>
<th>( H_0: \pi_2 = 0 )</th>
<th>( H_0: \pi_3 = 0 )</th>
<th>( H_0: \pi_1 = \pi_2 = \pi_3 = \pi_4 = 0 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Y_t</td>
<td>-0.7522 0.0127</td>
<td>-3.16458 0.0031</td>
<td>-0.92844 0.0035</td>
<td>-4.39898 0.0001</td>
</tr>
<tr>
<td>M2_t</td>
<td>-0.6181 0.0054</td>
<td>-1.7374 0.0906</td>
<td>-2.47449 0.0181</td>
<td>-3.95891 0.0003</td>
</tr>
<tr>
<td>ER_t</td>
<td>-0.41651 0.0126</td>
<td>-5.4178 0.0000</td>
<td>-3.05843 0.0041</td>
<td>-4.80454 0.000</td>
</tr>
<tr>
<td>EP_t</td>
<td>-2.21888 0.0327</td>
<td>-4.91828 0.000</td>
<td>-2.5907 0.0136</td>
<td>-3.85404 0.0004</td>
</tr>
<tr>
<td>FI_t</td>
<td>-0.89677 0.0063</td>
<td>-0.66739 0.0092</td>
<td>-1.26156 0.0216</td>
<td>0.46706 0.6435</td>
</tr>
<tr>
<td>FE_t</td>
<td>-1.55547 0.0083</td>
<td>-4.41476 0.0001</td>
<td>-1.04707 0.0019</td>
<td>-2.77022 0.0087</td>
</tr>
<tr>
<td>FP_t</td>
<td>-0.82130 0.0145</td>
<td>-3.5167 0.000</td>
<td>-2.05743 0.0041</td>
<td>-3.42030 0.000</td>
</tr>
</tbody>
</table>

From table 4.3 it is quite evident that all the null hypotheses \( H_0: \pi_1 = 0 \), \( H_0: \pi_2 = 0 \), \( H_0: \pi_3 = 0 \), \( H_0: \pi_1 = \pi_2 = \pi_3 = \pi_4 = 0 \) are non-significant and seasonal as well as non-seasonal unit roots are present in the time series data. So our data needs to be adjusted for seasonality.

4.3 Study of Long run and Causal Relationships Between Seasonally Unadjusted Variables

4.3.1 Granger Causality

The Granger causality test has been applied on the variables in the model and the results are presented in the table 4.4. The results presented here show that two-way causality does not exists between food prices and any other variable included in the model.

It is evident from the results that food prices granger cause money supply, energy prices and wheat support price. Whereas exchange rate and value added in agriculture granger cause food prices.
As per the definition of causality it can be interpreted from the results that exchange rate and value added in agriculture help predict food prices which is quite obvious for the case of value added in agriculture, but nexus between exchange rate and food prices can be true in the case of imported commodities only.

**Table 4.4: Granger Causality Tests**

<table>
<thead>
<tr>
<th>Null Hypothesis</th>
<th>F-Statistic</th>
<th>Prob.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exchange Rate does not Granger Cause Food Prices</td>
<td>1.32373</td>
<td>0.2561*</td>
</tr>
<tr>
<td>Food Prices does not Granger Cause Exchange Rate</td>
<td>5.85834</td>
<td>0.0197</td>
</tr>
<tr>
<td>Energy Prices does not Granger Cause Food Prices</td>
<td>18.7815</td>
<td>8.00E-05</td>
</tr>
<tr>
<td>Food Prices does not Granger Cause Energy Prices</td>
<td>1.50039</td>
<td>0.2271*</td>
</tr>
<tr>
<td>Money Supply does not Granger Cause Food Prices</td>
<td>9.79483</td>
<td>0.0031</td>
</tr>
<tr>
<td>Food Prices does not Granger Cause Money Supply</td>
<td>1.32033</td>
<td>0.2567*</td>
</tr>
<tr>
<td>Food Exports does not Granger Cause Food Prices</td>
<td>7.02447</td>
<td>0.0111</td>
</tr>
<tr>
<td>Food Prices does not Granger Cause Food Exports</td>
<td>12.7384</td>
<td>0.0009</td>
</tr>
<tr>
<td>Food Imports does not Granger Cause Food Prices</td>
<td>4.09871</td>
<td>0.049</td>
</tr>
<tr>
<td>Food Prices does not Granger Cause Food Imports</td>
<td>10.2293</td>
<td>0.0026</td>
</tr>
<tr>
<td>Value Added in Agriculture does not Granger Cause Prices</td>
<td>0.65568</td>
<td>0.4224*</td>
</tr>
<tr>
<td>Food Prices does not Granger Cause Value Added in Agriculture</td>
<td>4.96163</td>
<td>0.0311</td>
</tr>
<tr>
<td>Wheat Support Price does not Granger Cause Food Prices</td>
<td>22.6847</td>
<td>2.00E-05</td>
</tr>
<tr>
<td>Food Prices does not Granger Cause Wheat Support Price</td>
<td>1.04655</td>
<td>0.3119*</td>
</tr>
<tr>
<td>Fertilizer Prices does not Granger Cause Food Prices</td>
<td>42.2125</td>
<td>6.00E-08</td>
</tr>
<tr>
<td>Food Prices does not Granger Cause Fertilizer Prices</td>
<td>4.83534</td>
<td>0.0332</td>
</tr>
</tbody>
</table>

### 4.3.2 Johansen Cointegration Test

All the variables have same order of integration, therefore Johansen’s co integration can be applied to find the long-run relationship of food price inflation, money supply, wheat support prices, food exports, food imports, fertilizer prices, value added in agriculture, energy prices and exchange rate.

The results of Johansen’s co integration test have been reported in table 4.5a and 4.5b:

**Table 4.5a Unrestricted Co integration Rank Test**

<table>
<thead>
<tr>
<th>(Trace)</th>
<th>Hypothesized</th>
<th>Trace</th>
<th>0.05</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. of CE(s)</td>
<td>Eigenvalue</td>
<td>Statistic</td>
<td>Critical Value</td>
</tr>
<tr>
<td>None *</td>
<td>0.937858</td>
<td>421.1771</td>
<td>179.5098</td>
</tr>
<tr>
<td>At most 1 *</td>
<td>0.8875</td>
<td>298.9305</td>
<td>143.6691</td>
</tr>
<tr>
<td>At most 2 *</td>
<td>0.81204</td>
<td>202.7993</td>
<td>111.7805</td>
</tr>
<tr>
<td>At most 3 *</td>
<td>0.66609</td>
<td>129.2521</td>
<td>83.93712</td>
</tr>
<tr>
<td>At most 4 *</td>
<td>0.561189</td>
<td>80.9892</td>
<td>60.06141</td>
</tr>
<tr>
<td>At most 5 *</td>
<td>0.472071</td>
<td>44.74704</td>
<td>40.17493</td>
</tr>
<tr>
<td>At most 6</td>
<td>0.240656</td>
<td>16.64014</td>
<td>24.27596</td>
</tr>
<tr>
<td>At most 7</td>
<td>0.097551</td>
<td>4.526903</td>
<td>12.3209</td>
</tr>
<tr>
<td>At most 8</td>
<td>0.00024</td>
<td>0.010582</td>
<td>4.129906</td>
</tr>
</tbody>
</table>
Table 4.5b: Unrestricted Co-integration Rank Test
(Maximum Eigenvalue)

<table>
<thead>
<tr>
<th>Hypothesized</th>
<th>No. of CE(s)</th>
<th>Eigenvalue</th>
<th>Max-Eigen Statistic</th>
<th>Critical Value</th>
<th>Prob.**</th>
</tr>
</thead>
<tbody>
<tr>
<td>None *</td>
<td></td>
<td>0.937858</td>
<td>122.2466</td>
<td>54.96577</td>
<td>0.000</td>
</tr>
<tr>
<td>At most 1 *</td>
<td></td>
<td>0.8875</td>
<td>96.13121</td>
<td>48.8772</td>
<td>0.000</td>
</tr>
<tr>
<td>At most 2 *</td>
<td></td>
<td>0.81204</td>
<td>73.54722</td>
<td>42.77219</td>
<td>0.000</td>
</tr>
<tr>
<td>At most 3 *</td>
<td></td>
<td>0.66609</td>
<td>48.26289</td>
<td>36.63019</td>
<td>0.0015</td>
</tr>
<tr>
<td>At most 4 *</td>
<td></td>
<td>0.561189</td>
<td>36.24216</td>
<td>30.43961</td>
<td>0.0085</td>
</tr>
<tr>
<td>At most 5 *</td>
<td></td>
<td>0.472071</td>
<td>28.1069</td>
<td>24.15921</td>
<td>0.0139</td>
</tr>
<tr>
<td>At most 6</td>
<td></td>
<td>0.240656</td>
<td>12.11323</td>
<td>17.7973</td>
<td>0.2906</td>
</tr>
<tr>
<td>At most 7</td>
<td></td>
<td>0.097551</td>
<td>4.516321</td>
<td>11.2248</td>
<td>0.548</td>
</tr>
<tr>
<td>At most 8</td>
<td></td>
<td>0.00024</td>
<td>0.010582</td>
<td>4.129906</td>
<td>0.9332</td>
</tr>
</tbody>
</table>

Both the Maximum Eigen Statistics \( \lambda_{\text{max}} \) and Trace Statistics \( \lambda_{\text{trace}} \) are used to find co-integration and the number of co-integrating vectors. Both statistics confirm the existence of co-integration and same number (six) of co-integrating vectors.

4.3.3 Estimation of the Model For Food Inflation

The following table presents the results for the estimated model of food inflation using ordinary least squares procedure.

<table>
<thead>
<tr>
<th>Variables</th>
<th>Coefficient</th>
<th>StDev</th>
<th>t-statistic</th>
<th>prob.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>3.18</td>
<td>25.62</td>
<td>0.12</td>
<td>0.902*</td>
</tr>
<tr>
<td>( Y_{t-1} )</td>
<td>1.12028</td>
<td>0.06506</td>
<td>17.22</td>
<td>0.000</td>
</tr>
<tr>
<td>( EP_t )</td>
<td>0.23734</td>
<td>0.04107</td>
<td>5.78</td>
<td>0.000</td>
</tr>
<tr>
<td>( FP_t )</td>
<td>-0.03205</td>
<td>0.03351</td>
<td>-0.96</td>
<td>0.345*</td>
</tr>
<tr>
<td>( ER_t )</td>
<td>-0.1684</td>
<td>0.1591</td>
<td>-1.06</td>
<td>0.297*</td>
</tr>
<tr>
<td>( M2_t )</td>
<td>-4.7E-06</td>
<td>1.91E-06</td>
<td>-2.46</td>
<td>0.019*</td>
</tr>
<tr>
<td>( WSP_t )</td>
<td>0.01762</td>
<td>0.0116</td>
<td>1.52</td>
<td>0.137*</td>
</tr>
<tr>
<td>( VA_t )</td>
<td>0.2073</td>
<td>0.3649</td>
<td>0.57</td>
<td>0.573*</td>
</tr>
<tr>
<td>( FI_t )</td>
<td>-5.4E-06</td>
<td>6.46E-06</td>
<td>-0.83</td>
<td>0.409*</td>
</tr>
<tr>
<td>( FE_t )</td>
<td>-1.6E-05</td>
<td>1.41E-05</td>
<td>-1.16</td>
<td>0.253*</td>
</tr>
</tbody>
</table>

Note:*Represents the result is significant.

From the Table 4.6 we can see that in the long run lag of food prices i.e., adaptive expectations and energy prices are not significant whereas all other variables are significant and have considerable influence in determining food price inflation. As expected Exchange Rate (\( ER_t \)) has a negative coefficient and depreciation of 0.1684 in the value of rupee results in 1 rupee increase in food prices. Food imports and Food Exports also have negative
Determinants of food inflation in Pakistan & effects…

relationship with food prices whereas fertilizer prices, value added in agriculture and wheat support prices all have positive relationship with food prices.

Hence the estimated model for food prices is:

\[
\hat{Y}_t = 3.2 + 1.12 Y_{t-1} + 0.237 EP_t - 0.0320 FP_t - 0.168 ER_t - 0.000005 M_t^2 - 0.0176 WSP_t + 0.207 VA_t - 0.000005 FI_t - 0.000016 FE_t
\]  

(4.1)

Model diagnostics for the model of food prices are given below:

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>SS</th>
<th>MS</th>
<th>F-statistic</th>
<th>Prob.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regression</td>
<td>9</td>
<td>181079</td>
<td>20120</td>
<td>2626.85</td>
<td>0.000</td>
</tr>
<tr>
<td>Error</td>
<td>38</td>
<td>291</td>
<td>8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>47</td>
<td>181370</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

It is evident from Table 4.7 that the regression 4.1 is highly significant as p-value for the null hypothesis of insignificant regression is very small and it is rejected.

Coefficient of determination and adjusted coefficient of determination for the estimated model are quite high that is an indication that the model is a good fit.

\[
R^2 = 99.8\% 
\]

\[
\overline{R^2} = 99.8\%
\]

4.3.4 Forecasting Food Inflation

Forecasts for food inflation based on the model 4.1 are given below along with graph for forecasts.

<table>
<thead>
<tr>
<th>Horizon</th>
<th>Forecast</th>
<th>SE</th>
<th>Actual</th>
<th>Error</th>
<th>t-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>2009-2</td>
<td>130.946</td>
<td>34.92</td>
<td>221</td>
<td>90.0541</td>
<td>2.579</td>
</tr>
<tr>
<td>2009-3</td>
<td>130.946</td>
<td>34.92</td>
<td>233</td>
<td>102.054</td>
<td>2.922</td>
</tr>
<tr>
<td>2009-4</td>
<td>130.946</td>
<td>34.92</td>
<td>240</td>
<td>109.054</td>
<td>3.123</td>
</tr>
<tr>
<td>2010-1</td>
<td>130.946</td>
<td>34.92</td>
<td>245</td>
<td>114.054</td>
<td>3.266</td>
</tr>
<tr>
<td>2010-2</td>
<td>130.946</td>
<td>34.92</td>
<td>253</td>
<td>122.054</td>
<td>3.495</td>
</tr>
<tr>
<td>2010-3</td>
<td>130.946</td>
<td>34.92</td>
<td>260</td>
<td>129.054</td>
<td>3.696</td>
</tr>
<tr>
<td>2010-4</td>
<td>130.946</td>
<td>34.92</td>
<td>263</td>
<td>132.054</td>
<td>3.782</td>
</tr>
<tr>
<td>2011-1</td>
<td>130.946</td>
<td>34.92</td>
<td>268</td>
<td>137.054</td>
<td>3.925</td>
</tr>
<tr>
<td>2011-2</td>
<td>130.946</td>
<td>34.92</td>
<td>270</td>
<td>139.054</td>
<td>3.982</td>
</tr>
<tr>
<td>2011-3</td>
<td>130.946</td>
<td>34.92</td>
<td>279</td>
<td>148.054</td>
<td>4.24</td>
</tr>
</tbody>
</table>

mean(Error) 122.25 RMSE = 123.47
SD(Error) 17.308 MAPE = 48.034
4.4 Study of Long-run & Causal Relationships
After Seasonal Adjustment of the Data

In section 4.3 sufficient evidence was found about the existence of seasonal variation in the time series data used for the purposes of estimation, so it becomes obligatory to perform seasonal adjustment on the data before proceeding further. So in order to adjust for seasonality software package named Demetra was employed (details given in Chapter 4).

4.4.1 Causality Tests
After making seasonal adjustment of the data Granger Causality test was again performed to know the causal links between the variables. The results obtained are different and more pragmatic than the ones obtained without seasonal adjustment of the data. Results are given in the table 4.9.

Table 4.9: Test for Granger Causality on Seasonally Adjusted Data

<table>
<thead>
<tr>
<th>Null Hypothesis</th>
<th>F-Statistic</th>
<th>Prob.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Food Prices does not Granger Cause Energy Prices</td>
<td>1.33451</td>
<td>0.281*</td>
</tr>
<tr>
<td>Energy Prices does not Granger Cause Food Prices</td>
<td>3.0052</td>
<td>0.0194</td>
</tr>
<tr>
<td>Food Prices does not Granger Cause Exchange Rate</td>
<td>3.26693</td>
<td>0.0131</td>
</tr>
<tr>
<td>Exchange Rate does not Granger Cause Food Prices</td>
<td>0.99317</td>
<td>0.4758*</td>
</tr>
<tr>
<td>Food Prices does not Granger Cause Export Prices</td>
<td>0.91265</td>
<td>0.5337*</td>
</tr>
<tr>
<td>Export Prices does not Granger Cause Food Prices</td>
<td>5.58088</td>
<td>0.0007</td>
</tr>
<tr>
<td>Fertilizer Prices does not Granger Cause Food Prices</td>
<td>6.47595</td>
<td>0.0003</td>
</tr>
<tr>
<td>Food Prices does not Granger Cause Fertilizer Prices</td>
<td>1.08789</td>
<td>0.4133*</td>
</tr>
<tr>
<td>Import Prices does not Granger Cause Food Prices</td>
<td>4.15332</td>
<td>0.0038</td>
</tr>
<tr>
<td>Food Prices does not Granger Cause Import Prices</td>
<td>0.59701</td>
<td>0.7847*</td>
</tr>
<tr>
<td>Money Supply does not Granger Cause Food Prices</td>
<td>5.05099</td>
<td>0.0012</td>
</tr>
<tr>
<td>Food Prices does not Granger Cause Money Supply</td>
<td>1.86742</td>
<td>0.1174*</td>
</tr>
<tr>
<td>VA does not Granger Cause Food Prices</td>
<td>0.6212</td>
<td>0.7655*</td>
</tr>
<tr>
<td>Food Prices does not Granger Cause VA</td>
<td>2.15499</td>
<td>0.0733*</td>
</tr>
<tr>
<td>WSP does not Granger Cause Food Prices</td>
<td>1.14758</td>
<td>0.3773*</td>
</tr>
<tr>
<td>Food Prices does not Granger Cause WSP</td>
<td>0.94622</td>
<td>0.5091*</td>
</tr>
</tbody>
</table>
As manifested by the results we can notice two-way causality exists between value added in agriculture and food prices. Similarly Wheat Support Prices and Food Prices also have reverse causal links and both play equal part in determining each other.

Besides two-way causal nexus we also notice that food prices granger cause energy prices and food exports. In addition to this it was found that Exchange rate, Fertilizer prices, Food imports and Money supply all granger cause food prices.

4.4.2 Johansen Cointegration Test

To study the existence of long-run relationships between the variables that have met with seasonality adjustment Johansen’s test has been conducted as earlier and the results are presented in the table below:

<table>
<thead>
<tr>
<th>Hypothesized</th>
<th>Trace</th>
<th>0.05</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. of CE(s)</td>
<td>Eigenvalue</td>
<td>Statistic</td>
</tr>
<tr>
<td>None *</td>
<td>0.879442</td>
<td>340.6323</td>
</tr>
<tr>
<td>At most 1 *</td>
<td>0.797446</td>
<td>243.3136</td>
</tr>
<tr>
<td>At most 2 *</td>
<td>0.719993</td>
<td>169.863</td>
</tr>
<tr>
<td>At most 3 *</td>
<td>0.595189</td>
<td>111.3078</td>
</tr>
<tr>
<td>At most 4</td>
<td>0.408247</td>
<td>69.70843</td>
</tr>
<tr>
<td>At most 5</td>
<td>0.376276</td>
<td>45.57378</td>
</tr>
<tr>
<td>At most 6</td>
<td>0.272322</td>
<td>23.8596</td>
</tr>
<tr>
<td>At most 7</td>
<td>0.124704</td>
<td>9.236327</td>
</tr>
<tr>
<td>At most 8</td>
<td>0.065362</td>
<td>3.109428</td>
</tr>
</tbody>
</table>

Both the Maximum Eigen Statistics $\lambda_{\text{max}}$ and Trace Statistics $\lambda_{\text{trace}}$ are used to find Co-integration and the number of co-integrating vectors. Both statistics confirm the existence of co-integration and same number (four) of co-integrating vectors.

4.4.3 Estimation of the Model For Food Inflation

The following table presents the results for the estimated model of food inflation using the data that has been adjusted for seasonality using ordinary least squares procedure.
Table 4.11: Long-run Coefficients
(Dependent Variable is Food Prices $Y_t$)

<table>
<thead>
<tr>
<th>Variables</th>
<th>Coefficient</th>
<th>StDev</th>
<th>t-statistic</th>
<th>prob.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>-15.08</td>
<td>25.46</td>
<td>-0.59</td>
<td>0.557*</td>
</tr>
<tr>
<td>$Y_{t-1}$</td>
<td>0.9488</td>
<td>0.1001</td>
<td>9.47</td>
<td>0.052*</td>
</tr>
<tr>
<td>$EP_t$</td>
<td>0.17719</td>
<td>0.05559</td>
<td>3.19</td>
<td>0.523*</td>
</tr>
<tr>
<td>$FP_t$</td>
<td>0.04978</td>
<td>0.03897</td>
<td>1.28</td>
<td>0.209*</td>
</tr>
<tr>
<td>$ER_t$</td>
<td>-0.0116</td>
<td>0.1525</td>
<td>-0.08</td>
<td>0.94*</td>
</tr>
<tr>
<td>$M2_t$</td>
<td>-4.8E-06</td>
<td>1.69E-06</td>
<td>-2.81</td>
<td>0.008*</td>
</tr>
<tr>
<td>$WSP_t$</td>
<td>0.003649</td>
<td>0.009616</td>
<td>0.38</td>
<td>0.706*</td>
</tr>
<tr>
<td>$VA_t$</td>
<td>0.104</td>
<td>0.3304</td>
<td>0.31</td>
<td>0.755*</td>
</tr>
<tr>
<td>$FI_t$</td>
<td>-1.6E-05</td>
<td>1.08E-05</td>
<td>-1.52</td>
<td>0.137*</td>
</tr>
<tr>
<td>$FE_t$</td>
<td>-7E-08</td>
<td>2.21E-05</td>
<td>0</td>
<td>0.997*</td>
</tr>
</tbody>
</table>

Note:*Represents the result is significant.

From the Table 4.11 we can see contrary to the results obtained when the data was not adjusted for seasonality now the lag of food prices i.e., adaptive expectations and energy prices are also significant along with all other variables and they also contribute significantly in determining food price inflation. Exchange Rate ($ER_t$), Money Supply ($M2_t$), Food Imports ($FI_t$) and Food Exports ($FE_t$) all have negative influence on the food price inflation in Pakistan whereas Fertilizer Prices ($FP_t$), value added in agriculture ($VA_t$) and wheat support prices ($WSP_t$) all have positive relationship with food prices.

Hence the estimated model for food prices is:

$$
\hat{Y}_t = -15.1 + 0.949 Y_{t-1} + 0.177 EP_t - 0.0498 FP_t - 0.012 ER_t - 0.000005 M2_t - 0.00365 WSP_t + 0.104 VA_t - 0.000016 FI_t - 0.000000 FE_t
$$

Model diagnostics for the model of food prices are given below:

Table: 4.12 ANOVA Test for the significance of Regression 4.2

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>SS</th>
<th>MS</th>
<th>F-statistic</th>
<th>Prob.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regression</td>
<td>9</td>
<td>180182</td>
<td>20020</td>
<td>3638.6</td>
<td>0.000</td>
</tr>
<tr>
<td>Error</td>
<td>38</td>
<td>209</td>
<td>6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>47</td>
<td>180391</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

It is evident from Table 4.12 that the regression 4.2 is highly significant as p-value for the null hypothesis of insignificant regression is very small and it is rejected.

Coefficient of determination and adjusted coefficient of determination for the estimated model are quite high that is an indication that the model is a good fit.

$$
R^2 = 99.98\% \quad \bar{R}^2 = 99.98\%
$$
4.4.4 Forecasting Food Inflation

Forecasts for food inflation based on the model 4.2 are given below along with graph for forecasts.

Table 4.13: Dynamic (ex ante) forecasts for Yt (SE based on error variance)

<table>
<thead>
<tr>
<th>Horizon</th>
<th>Forecast</th>
<th>SE</th>
<th>Actual</th>
<th>Error</th>
<th>t-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>2009-2</td>
<td>130.987</td>
<td>34.93</td>
<td>222.015</td>
<td>91.028</td>
<td>2.606</td>
</tr>
<tr>
<td>2009-3</td>
<td>130.987</td>
<td>34.93</td>
<td>230.502</td>
<td>99.515</td>
<td>2.849</td>
</tr>
<tr>
<td>2009-4</td>
<td>130.987</td>
<td>34.93</td>
<td>237.809</td>
<td>106.822</td>
<td>3.058</td>
</tr>
<tr>
<td>2010-1</td>
<td>130.987</td>
<td>34.93</td>
<td>248.845</td>
<td>117.858</td>
<td>3.374</td>
</tr>
<tr>
<td>2010-2</td>
<td>130.987</td>
<td>34.93</td>
<td>254.137</td>
<td>123.15</td>
<td>3.526</td>
</tr>
<tr>
<td>2010-3</td>
<td>130.987</td>
<td>34.93</td>
<td>257.35</td>
<td>126.363</td>
<td>3.618</td>
</tr>
<tr>
<td>2010-4</td>
<td>130.987</td>
<td>34.93</td>
<td>260.442</td>
<td>129.456</td>
<td>3.706</td>
</tr>
<tr>
<td>2011-1</td>
<td>130.987</td>
<td>34.93</td>
<td>272.236</td>
<td>141.25</td>
<td>4.044</td>
</tr>
<tr>
<td>2011-2</td>
<td>130.987</td>
<td>34.93</td>
<td>271.242</td>
<td>140.255</td>
<td>4.015</td>
</tr>
<tr>
<td>2011-3</td>
<td>130.987</td>
<td>34.93</td>
<td>276.154</td>
<td>145.167</td>
<td>4.156</td>
</tr>
<tr>
<td>mean(Error)</td>
<td>122.09</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SD(Error)</td>
<td>17.403</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 4.2

CONCLUDING REMARKS

In the recent years, food price inflation has risen very sharply at global level. It has increased the living cost of households especially in developing countries like Pakistan which results in malnutrition and, therefore, productivity losses. It hurts the poor more because they spend more than half of their budget on food. In Pakistan, food inflation remained 9.9% on average during the study period (1972-2008), some time as high as 34.7% in 1974 and 26.6% in 2008-09. Food inflation drives people to demand more in wages and this finally pushes up non-food inflation. So, in fighting inflation, food inflation should be the main thrust for the government. According to the structuralist school of thought, inflation arises from two bottlenecks, namely agricultural bottleneck and foreign exchange bottleneck. The economists of this school of thought believes that in developing countries there are structural problems enshrined in the agricultural system which allows food price to rise which pushes up the cost of living. This leads to an increase in the wages and salaries with the presence of expected inflation. When this
rising demand from an increasing population outpaces the growth in food supply, food price increases.

Our demand is not outpacing the growth in food supply, but still we are faced with this nagging problem. The market structure may provide the answer. With the absence of strict market regulation, there is anarchy in the market. Just consider the price of vegetables which you will find at very different levels in different markets. Here strengthening of the Trading Corporation of Pakistan can go a long way to mitigating this problem. Again, if a system could be introduced where farmers will directly sell to the end-market, the middlemen’s share in food items will greatly reduce and hence will work to reduce inflation. These are structural problems and what about our government’s policies taken for the last couple of months. The bank borrowing of the government for meeting its fiscal expenditure has always been growing at a rapid pace. Now as borrowing from the central bank normally means printing money, this will result in the expansion of this amount of money. This will directly result into excess liquidity problem and has all the potential to increase the price level higher.

In conducting this research Time series data from 2000-Q1 to 2011-Q4 of relevant variables was used for empirical analysis. First of all, stationary of time series was checked by using Augmented Dickey-Fuller (ADF) unit root test. All variables were integrated of order one I(1) as they became stationary at their first differences at 5% level of significance. After that Seasonal unit root test using HEGY procedure was conducted and it indicated the presence of seasonal unit roots that can’t be accounted for by introducing the seasonal dummies in the model rather the data needed to be seasonally adjusted. Software package Demetra was used to adjust the time series data on the relevant variables for seasonal factor. As the variables had same order of integration, therefore Johansen co-integration was applied to find the long run relationships. Maximum Eigen Statistics and Trace Statistics were used. Both statistics confirmed the existence of co-integration and same number (six) of co-integrating vectors for the unadjusted data and four co integrating vectors for the data that was adjusted for seasonality. Besides Johansen co integration, causal links between the relevant variables were also checked to find out which variables help predict food prices in the future.

The results revealed that both demand and supply side factors determined food price inflation in Pakistan. However, on the basis of empirical results we may conclude that food price inflation is not a monetary phenomenon in Pakistan (money supply growth is statistically insignificant). While the supply side factor or structural factors have dominant role in determining the food prices. Findings of the study show that prices of energy (petroleum products, gas and electricity), exchange rate, value added in agriculture and, government’s monetary policy and fiscal policy have dominant role in determining high inflation in Pakistan. Support prices are the second major source of food price inflation in Pakistan. Government should pursue a moderate policy in raising support prices. Alternative to support price policy, government may provide subsidies on inputs as on fertilizers, pesticides, diesel and electricity. Government should also encourage and support farmers to adopt modern technology for higher production with lower production cost. Sufficient credit facilities should be provided through formal and informal channel. Government should take measures to improve infrastructure, agriculture markets and land ownership system. Modern technology should be introduced
to improve the production of food grains, meat, poultry and dairy products. According to our analysis, imports of food items are also inflationary because of higher prices of food item at global level and exchange rate depreciation. As a policy measure, we need to exploit our unrealized yield potential in production of food items as God has gifted us with all necessary resources. This study reveals that food exports affect food price inflation positively. Government should ban the exports of food items until they are over and above the domestic needs. For price stability in the country, buffer stocks of essential food items like wheat, sugar and pulses should be maintained. There should be maximum control on smuggling of wheat, rice and live stock to neighboring countries.

REFERENCES

BIOMETRICS SUPPORT AND AGRICULTURAL STATISTICS SYSTEM: TOOLS OF PRODUCTIVITY INCREASE

Osman M. Siraj\textsuperscript{1} and M. Singh\textsuperscript{2}

\textsuperscript{1} Experimental Design and Analysis Unit, Agricultural Research Corporation
Wad Medani, Sudan. Email: sirajstat@yahoo.com

\textsuperscript{2} International Center for Agricultural Research in the Dry Areas (ICARDA)
Aleppo, Syria. Email: m.singh@cgiar.org

ABSTRACT

Biometrics training and development of agricultural statistics systems is an exciting scientific field in which statistical practice and methodology serve as tools for operations research in process of productivity increase. This paper addresses the role of Biometric support in agricultural research and agricultural production system for development. Data was obtained from the website http://www.cbssyr.org, the Central Bureau of Statistics (CBS), Syria, on area, and production & yield of total wheat by governorate and at the country level during 2000-2009. The data's example was analyzed using descriptive statistics, test-t, and regression analysis to study the production pattern. The results showed that there was a significant difference (P<0.01) between irrigated and non-irrigated area for production and yield of wheat crops. Linear regression of yield on area under irrigation explained 40% of variation. For non-irrigated condition, there was no statistically significant effect of area. Overall, the relation between area and yield was not significant, also statistical control chart was used for monitoring the yield.

Biometric techniques and tools are used during planning of experiments to generate experimental designs. Once the data are available, statistical analysis is performed, results are interpreted, conclusions are drawn and presentation of results made on the lines of the objectives set. The Agricultural Statistics system in general and Crop Statistics system in particular is an instrument used in agricultural policy analysis. The recommendation here is mode to highlight e the role of team work between agricultural scientists, agricultural statistician and biometrician to address challenges and to contribute to the change in development of agricultural research, practices and policies.

KEYWORDS

Agricultural Research; Biometrics; Agricultural Statistics.

1. INTRODUCTION

In February 2009, the United Nation's Statistical Commission endorsed the global strategic framework for improving agricultural statistics. In 1997, Arab organization and the agricultural development demonstrated weaknesses of knowledge in the agricultural statistics system in Arab countries and presented a proposal on capacity building in agricultural statistics to address the problem of insufficient number of agricultural
information. Food and Agriculture Organization of the United Nations organized a plan for the world census of agriculture 2010 programme: which includes a modular approach a round of agricultural censuses, to be undertaken between 2006 and 2015 into the ongoing system of agricultural statistics (Siraj et al. 2011).

Agricultural statistics such as production, environment, social-economy and rural development are a rich and important source of information to apply statistical analysis methods, collaborate and statistics for measuring progress in our society, (Ankush et al. 2010). An agricultural statistics strategy is a very important tool in guiding the collection, processing, analysis and dissemination of a country’s agricultural statistics (Edmund, 2010). The primary purpose of agricultural statistics is to provide timely, accurate and useful statistics to estimate crop production, livestock, production practices, farm economics, etc. (Michael et al. 2010). The development of newer statistical applications for data collection, policy planning and maintenance of information systems on various components of economic interest are very important (Bhatia 2010).

An innovation agricultural production system has played a significant role in the increased productivity and sustainability (Vogel, 1995). There are challenges ahead to understand the different data on population growth, food demand, water availability, climate change, and global growth systems thinking, where we are today are even more critical to meet the challenges of the future, the study suggest to build a agricultural statistics production systems as new approach and body generated from national agricultural statistics (NAS) with biometricians researcher offices. Major domains of agricultural statistics include statistics related to production, land use, prices, fertilizers, employment, credit and trade of agriculture produces (Ankush et al. 2010).

2. JUSTIFICATION OF THE STUDY

The conduct of the crop area survey is not simple. The crop area should not be the physical area measured on the slope but its projection on the horizontal plane especially in the dry areas. In addition to difficulties associated with agricultural statistics system in Arab countries and measuring the agricultural standards, there are difficulties in determining the share of specific crops in presence of mixed or associated cropping. The problem of estimating crop areas in these situations gets more and more complicated as the number of crops in the mixture or in association increases and especially when the proportions of the different crops in the mixture vary from field to field. Moreover, the crops may have different periods of sowing, planting and harvesting.

Challenges consider is a consistency for adopting of modern statistics application in agricultural statistics system. Gives special focus to the application and value of the data standards especially as it may support the construction of indicators for productivity increase. Biometry support assesses and explores the potential for improving policy-relevant indicators and in empirical analysis of policies for farm and rural households, natural resources, and regional economic development. Estimates on production are derived from the estimates on area and the yield, and then the biometricians must be involved as part of a team work because he play a significant role in estimation or inference statistical of yield production.
So more than four decades of statistical work in many countries, the need to build statistical capacity and systems is as great now as it has ever been. Fortunately, a number of initiatives have recently been put in place to improve statistics in general and agricultural statistics in particular.

Objectives
This paper addresses the role of biometrics support in agricultural research and agricultural statistics system with focusing on agricultural statistics production because these form the backbone of the database required for estimation of yield.

3. AGRICULTURAL STATISTICS SYSTEM

The agricultural statistics system is very comprehensive and provides data on a wide range of topics such as crop area and production, land use, irrigation, land holdings, agricultural prices and market intelligence, livestock, fisheries, forestry, etc. Recent challenges have forced us to greater innovation in design, analysis and presentation of results on the work into agriculture and rural developments (Siraj, 2010). In order to enhance agricultural statistical integration we need of capacity building in different aspects such as, agricultural economics, Biometry and computer application (Barbara et al. 2008). The activities of agricultural statistics are responsible for statistical surveys to collect data/information in the field of agriculture and rural development to provide them to governors and policymakers such as crop survey, livestock survey, fishery survey, forestry survey, structural survey, and other surveys (Siraj, 2010).

4. ENVIRONMENT IS CHANGING AND BIOMETRICIAN NEEDS ARE GROWING

- The development of agricultural statistics is closely linked with the economic and social development of a society and its administrative structures. During the past decade, discussion about the information society and its future prospects has been ongoing among politicians, researchers and statisticians/biometricians. It is evident that the development of the information society causes a lot of changes in official statistics.
- In spite of the fact that the information society presents new challenges to agricultural statistics. Today, like in the past, wide and manifold supply of statistical information is necessary for the proper functioning in improvement of agricultural production. Furthermore, a high-quality statistical infrastructure enhances the ability to make change in means development.
- Briefly; there is no any development without linkages to statistics. Furthermore, follow-up and monitoring agricultural development based on statistical information.
- Statistics and statistical indicators have an important role in empirical studies and model building supporting policy making.

6. THE IMPORTANCE OF AGRICULTURAL STATISTICS

Agricultural statistics are important for a number of reasons (Jairo, 2010), to mention but the most important:
- Monitoring the food security of the nation;
- Informing policy making in the agricultural sector;
- Monitoring the progress towards achieving the millennium development goals MDGs linked to agriculture;
- Measuring the sector’s contribution to GDP;
- Measuring the performance of the sector;

7. THE FRAMEWORK OF AGRICULTURAL PRODUCTION STATISTICS SYSTEM

The following system of agricultural production statistics is designed by the author as results of previous studies of statistics system. The objectives and strategies of framework for agricultural production statistics system (APSS) are intended to mainstream sectorial statistical systems within the NAS. The sets of strategies have been organized by objectives in a framework that may be used when preparing an integrated system. There are four objectives with sets of corresponding strategies. The framework for APSS is described in the figure below.

![Fig. 1: Framework of the Agricultural Statistics production System](image)

7.1 Classifications and Standards

Classifications and standards allow comparisons to be readily made over time and across geographic areas, industries and other domains, the classifications and standards would be to:
• Using standard definitions for the subject-matter concepts, variables and classifications used, as well as the populations and statistical units to which they apply.
• Promoting of, and support for the use of statistical standards and classifications in official statistics produced by other agencies.
• Providing a designed classification for systematic ordering of criminal offences defined in the criminal laws.

7.2 Data Quality Frameworks
To develop and implement a quality framework to the objectives of the agricultural production function, there is need for a statistician function working in National Agricultural statistical production bodies. However, due to differences in the institutional environments the models developed differ somewhat with regard to their stakeholders, definitions of quality and scope at its most basic. Data quality framework is a tool for the assessment of data quality within an organization describe data environment, identify and analyze relevant data quality attributes in their current or future context, providing guidance for data quality improvement and solve data quality problems with their proactive management (Flavia et al. 2010).

7.3 Survey Design and Data Collection
The uses of design surveys and implement methods of collecting data will reduced the number of errors in our surveys and ensure the resulting statistics to support the main uses of the data (Cheung, 2010).

7.3.1 Survey design
• A guide to good survey design.
• Methods of statistics.
• Non-Sampling Error in Economic Surveys.
• Methodological standards for data collection.

7.3.2 Methodological standard for writing and constructing a questionnaire
• About methodological standards for data collection.
• Flowcharting specifications for electronic social questionnaires.
• Layout for self-administered paper questionnaires.
• Showcards for face-to-face surveys data analysis.

Agricultural production surveys are conducted following systematic design, collection, estimation and analytical methodologies. Agricultural production survey collected information from farmers, horticulturists, and foresters, and included selected farming practices, such as fertilizer application and cultivation practices (Siraj et al. 2011).

7.4 Data Analysis
Analytical methods in APSS include confidentiality techniques used for this purpose. Statistical analysis include
• Exploratory data analysis (EDA), an important part of any preliminary analysis of data.
• Time series analysis, mainly seasonal adjustment and trend estimation.
- Multivariate analysis, such as multivariate regression and principal component analysis.
- Longitudinal data analysis, which has become important with current development of longitudinal surveys.

### 7.5 Data Integration

- Data integration projects.
- Further technical information.
- Inland Revenue sourced tax data Available files.
- Data Integration Manual.

### 7.6 Data Integration

The role of data integration in helping to produce an effective official statistical system. Through integration it becomes possible to examine underlying relationships between various cross-sections of society, thus improving our knowledge and understanding about a particular subject.

### 8. BIOMETRICIANS

Biometricians advise researchers who are planning experiments, or they analyze data for trends and underlying relationships. The biometrician’s skills are important to quality control and assurance in research and operations management. In a large organization, a biometrics staff may support others in projects such as forest inventory design, agricultural research, and production process analysis. Biometricians may conduct their own research to improve statistical methods or the understanding of subjects like crop development or climate change (Dicks et al. 2009). Biometricians work for state agencies seeking staff support for research, planning, or information management; federal research organizations; private firms with large research programs; or large environmental consulting firms who want their skills in analyzing and evaluation (Janssen, 2009). Statistical agencies hire biometricians to support biological researchers to teach and do research on applied problems.

For over a century, increase yield have been one of the most highly used statistics methods (Coe, 2009). Biometry in recent years increase applied for investigating the relationship between on-increase yield trails. Nowadays statistical science as the main contributor to contemporary of agricultural developments and the apply biometry play a vital role for increasing the agricultural crop production and economic development.

The biometrician offers assistance in study design, data management and exploratory analysis, data modeling, analysis and interpretation, and survey methods.

### 9. BIOMETRIC SUPPORT

Biometry support in agriculture come in research, on-farm experiments, field surveys, social, health and environmental studies (Allan et al. 2009). Biometric support functions are provided most effectively when biometricians are in close contact with the scientists, with whom they work research support, increasing the efficiency and effectiveness of their use. The effect would be to:
• Increase the range of biometric and other research support skills available to research teams.
• Provide local and accessible support, allowing support staff to be active research team members.
• Increase the speed and effectiveness with which new methods and approaches are disseminated to all researchers in the region.
• Provide a base from which researchers in all projects in the region (whether at headquarters or not) can get support.
• Provide a body with the man-date, skills and resources to undertake effective and structured capacity building.
• Provide a body of sufficient weight that can attract funding and `buy in' from other organizations.
• Encourage interactions among biometricians and hence increase their capacity to undertake research on the development of new methods.

Biometrics training will implemented as important part of statistical research group for assessing and evaluating of the ASP activities for further research work (Siraj, 2009). For bringing overall improvements in the functioning of Agricultural Statistical Systems the training will be in two disciplines: 1) Training in the field of agricultural statistics and computer application 2) Training in biometry because agricultural research and practices depends on challenges such as weather, pests and uncertain markets which faces farmers and sensitive environmental concerns.

10. RESEARCH PAPERS

The working research papers is a series with analytical and methodological research on a variety of statistical topics linked with statistics production and related agricultural statistics databases on the usefulness, reliability, coverage and availability of the statistics produced and working on how to improve the official statistics system with integrated data from survey and administrative sources and agricultural understanding.

11. RESEARCH STATISTICS GROUP

The aim of the Research Biometry Group of the ASPS is to address specific methodological challenges faced collection of samples and analysis of agricultural statistics data. The proposal takes three components into account, following key technical components that need to be developed to increase agricultural productivity.

Regarding the first step focusing on the research areas which have the highest impact on the process of improvement of agricultural productive are to

• Improvement of estimation of crop area, yield and production, especially in the presence of mixed and/or repeated cropping, yield of root crops, small area estimation, edible forest products, etc.;
• Testing of integration of remote sensing into the production of agricultural statistics methodology for food balance sheets compilation, informal cross border trade data, food stocks, farm gate prices;
• Reconciling census data with current survey data;
• determination of user information needs for decision making;
• integrated agricultural survey methodology (master sampling frames and database); and
• Integration of administrative data for improving agricultural statistics.

12. WORK EXAMPLE

The purpose of this example is to illustrate statistical analysis application rather than using traditional method. In general, some of official statistics provide statistical analysis in term of annual report which may not lead to the right conclusion and statistical development to meet objectives and suitable solution and making right decision, in this example we present, t-test for comparing the area, production and yield under irrigated and non irrigated method as well as multiple regression analysis with considering the yield is independent variable and the area & production are independence variables. Time series analysis has been computed by using linear trend on total wheat yield.

Table 1:
Area, production & yield of total wheat during (2000-2009)

<table>
<thead>
<tr>
<th>Year</th>
<th>Area (ha)</th>
<th>Production (kg)</th>
<th>Yield (kg/ha)</th>
<th>Area (ha)</th>
<th>Production (kg)</th>
<th>Yield (kg/ha)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Irrigated</td>
<td>Non-Irrigated</td>
<td></td>
<td>Irrigated</td>
<td>Non-Irrigated</td>
<td></td>
</tr>
<tr>
<td>2000</td>
<td>694469</td>
<td>2396573</td>
<td>3451</td>
<td>984328</td>
<td>708916</td>
<td>720</td>
</tr>
<tr>
<td>2001</td>
<td>682786</td>
<td>2744203</td>
<td>4019</td>
<td>1000998</td>
<td>2000420</td>
<td>1998</td>
</tr>
<tr>
<td>2002</td>
<td>752524</td>
<td>3275222</td>
<td>4352</td>
<td>926826</td>
<td>1500220</td>
<td>1619</td>
</tr>
<tr>
<td>2003</td>
<td>814533</td>
<td>3421429</td>
<td>4200</td>
<td>981482</td>
<td>1491564</td>
<td>1520</td>
</tr>
<tr>
<td>2004</td>
<td>857463</td>
<td>3392660</td>
<td>3957</td>
<td>973763</td>
<td>1144799</td>
<td>1176</td>
</tr>
<tr>
<td>2005</td>
<td>855876</td>
<td>3471110</td>
<td>4056</td>
<td>1047950</td>
<td>1197636</td>
<td>1143</td>
</tr>
<tr>
<td>2006</td>
<td>810127</td>
<td>3567671</td>
<td>4404</td>
<td>976532</td>
<td>1363854</td>
<td>1397</td>
</tr>
<tr>
<td>2007</td>
<td>791358</td>
<td>3130010</td>
<td>3955</td>
<td>876374</td>
<td>911090</td>
<td>1040</td>
</tr>
<tr>
<td>2008</td>
<td>730334</td>
<td>1973906</td>
<td>2703</td>
<td>755657</td>
<td>165407</td>
<td>219</td>
</tr>
<tr>
<td>2009</td>
<td>656927</td>
<td>3037582</td>
<td>4624</td>
<td>780448</td>
<td>664202</td>
<td>851</td>
</tr>
</tbody>
</table>

Source: Syria Central Bureau of Statistics (http://www.cbssyr.org)

Table 2:
Statistical analysis of comparison of area, production, yields (kg/ha) between irrigated and non-irrigated method

<table>
<thead>
<tr>
<th>Items</th>
<th>Area (ha)</th>
<th>Production (kg)</th>
<th>Yield (Kg/ha)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Irrigated</td>
<td>Non-Irrigated</td>
<td>Irrigated</td>
</tr>
<tr>
<td>Mean</td>
<td>764640</td>
<td>930436</td>
<td>3041037</td>
</tr>
<tr>
<td>SD</td>
<td>72082</td>
<td>96770</td>
<td>520216</td>
</tr>
<tr>
<td>P-Value</td>
<td>0.0001</td>
<td>0.0001</td>
<td>0.0001</td>
</tr>
</tbody>
</table>

Table 2 shows the statistical analysis of comparison of area, production, yields (kg/ha) between irrigated and non-irrigated method. The descriptive statistics such as
mean and standard deviation (SD) p-value were computed, there was a highly statistical significant (p=0.0001) between area, production, yields under irrigated and non-irrigated. The irrigated method given highly total yields wheat (3972 kg/ha) comparing to non-irrigated method.

Table 3:
Simple regression analysis of total wheat yields under irrigated and non-irrigated systems

<table>
<thead>
<tr>
<th>Agricultural system</th>
<th>Predictor</th>
<th>Coefficient</th>
<th>SE Coefficient</th>
<th>T-Test</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Irrigated*</td>
<td>Constant</td>
<td>3347</td>
<td>2050</td>
<td>1.63</td>
<td>0.141</td>
</tr>
<tr>
<td></td>
<td>Area</td>
<td>0.00082</td>
<td>0.00267</td>
<td>0.31</td>
<td>0.767</td>
</tr>
<tr>
<td>Non-Irrigated**</td>
<td>Constant</td>
<td>-1912</td>
<td>1330</td>
<td>-1.44</td>
<td>0.188</td>
</tr>
<tr>
<td></td>
<td>Area</td>
<td>0.00331</td>
<td>0.00142</td>
<td>2.33</td>
<td>0.048</td>
</tr>
<tr>
<td>Over all***</td>
<td>Constant</td>
<td>316</td>
<td>1715</td>
<td>0.18</td>
<td>0.858</td>
</tr>
<tr>
<td></td>
<td>Area</td>
<td>0.00125</td>
<td>0.00101</td>
<td>1.24</td>
<td>0.25</td>
</tr>
</tbody>
</table>

*Yield (kg)= 3347 + 0.00082 Area (ha); R² = 1.2%
**Yield (kg)= -1912 + 0.00331 Area (ha); R² = 40.4%
****Yield (kg) = 316 + 0.00125 Area (ha); R² = 16.1%

where R² is variation explained by the Area (also called coefficient of determination).

Table 3 demonstrates that linear regression of total wheat yield versus area under irrigated was not statistical significant with 1.2% coefficient of determination but was statistical significant between yield and area under non-irrigated with (p=0.048) and 40% coefficient of determination.

![Linear Trend Model](image)

Fig. 2: Linear trend model of total wheat yield under irrigated method
Fig. 3: Linear trend model of total wheat yield under non-irrigated method

Figure 2 and 3 shows the trend model with original data, fitted trend line, and forecasts. There were no statistical significant of trend analysis or linear trend total yield of wheat of yield under irrigated or non-irrigated method. The fitted trend and three measures help on determined the: MAPE (mean absolute prediction error), MAD (mean absolute deviation), and MSD (mean squared deviation).

13. CONCLUSIONS

The agricultural statistics system must be designed to the actual situation in a country, or even a region, therefore agricultural statistics production as important tools to understand agriculture productivity.

Biometrics is an essential part in the agricultural research during planning, use of effective statistical design, and statistical analysis in depth, interpretation and presentation agricultural statistics system in general and crop statistics system or agricultural statistics production system in particular is an instrument in support of agricultural policy analysis.

Based on the results of the study the recommendations should is to creation cooperation team work between agricultural scientists, agricultural statistician and biometrician to meet challenges and making the change in development of agricultural research, practices and policies.

Agricultural institutions or agencies must promotion on agricultural statistics and certainly confirmed the need and importance of and census agricultural production statistics systems within framework of agricultural statistics systems with emphasized the role of biometry and biometrician.
Enhancing the environment of work, continuity of training courses to ensure capacity building and financial support to agricultural statistics program.

Governments, agricultural institutions, and policy makers must emphasized the importance of statistics in general and agricultural statistics in particular, and to implement the role of statisticians/biometrician by new approach and its applications in to push the sustainable development aims and take advantage of planning and making the right decisions about an agricultural development.

The future study has take place this framework is concerned with the development and establishment of agricultural statistics production system based on survey design, data quality framework and biometry support (Research statistics Group).

14. REFERENCES


ABSTRACT

Knowledge Elicitation out of an expert system is the major focus of studies in the domain of knowledge engineering. To establish knowledge base systems the role of knowledge engineering is critical therefore without proper knowledge elicitation, the reliability of knowledge base systems becomes questionable. Knowledge elicitation and its transfer to a knowledge base is not only a complex process but also needs significant categorization between implicit and explicit knowledge. This paper presents the important issues underlying the elicitation of explicit and implicit knowledge using artificial general intelligence and solution development on the basis of agency experience.

KEYWORDS

Knowledge elicitation; knowledge engineering; implicit knowledge; explicit knowledge; knowledge base systems; knowledge acquisition.

1. INTRODUCTION

Knowledge is expertise and skill required by a person through experience and education the theoretical and the practical understanding of a subject.

The process of knowledge elicitation for knowledge base system (KBS) is usually called bottleneck of their development. Inappropriate knowledge in the beginning of the development of KBS may makes all later on steps of elicitation wrong. So, how to gain adequate knowledge? There are many factors on which knowledge elicitation depends on. Most of them are primary psychological. Most of experts express their tacit (implicit) knowledge partially or sometimes not at all, that is the primary problem (Diaper, 1989b). Their implicit knowledge is gained through the year of experiences and it is built in their minds and they are using it without explicit thinking.

This kind of knowledge, experts never tried to express in words and it will become problem for elicitor to express such knowledge during the knowledge elicitation session. Elicitor must focus on all psychological aspects relevant for session and choose the best appropriate method for getting the desired knowledge. There are many elicitation techniques today. They are divided into groups:

1. Techniques for eliciting knowledge from Human Expert.
2. Techniques for eliciting knowledge from other sources.
Techniques for eliciting knowledge from human expert are: interviews focused interviews, construct elicitation, teach back, sorting task, laddering, “20 questions”, matrix generation, critiquing, protocols, role play and simulations.

Techniques for eliciting knowledge from other sources are: existing systems, physical or social environment and documents (Cordingley, 1989). All these techniques are used in elicitation; it depends upon the situation which one is to be selected.

Knowledge elicitation has its formal beginning in mid or late 1980s in the context of knowledge engineering for expert systems. Expert systems are computer programs that embody domain-specific knowledge and that perform decision making, problem solving, and design at levels typical of human experts, but not necessarily in exactly the same manner as human experts. Knowledge engineering is broadly defined here as the process of building knowledge-based systems or applications. These include expert systems, as well as intelligent systems, adaptive user-interfaces, and even knowledge-oriented selection and training devices.

The push for expert system in 70s and 80s was motivated by:

1. The technological capabilities
2. The growing specialization of the employees and cognitive complexity of jobs (Howell, 1989).
3. The interest in creating artificial intelligence in machines, and

Instead of relying on search strategies, this new form of machine intelligence was "knowledge-based" or powered by facts and rules. The realization that "knowledge is power," triggered a flurry of interest in knowledge, and particularly in its elicitation and representation (Feigenbaum, 1989).

Thus, knowledge elicitation is so demanding part in the development of KBS, finding the right technique is also important.

2. LITERATURE REVIEW

The knowledge elicitation problem causes the limitations when people are said to reveal, what they know in their area of specialization. Essentially, these are the main issue in extracting the useful knowledge from expert for developing the knowledge-based system. Knowledge engineering is an activity in which an expert has to elicit knowledge out of an expert. This knowledge then needs to be modeled such that it could be presented as set of rules in a rule base system or an expert system. Hsia et al describes knowledge as identifying and selecting contents from relevance, creating knowledge source catalog, capturing and discovering new knowledge (Hsia, 2006).

McGraw gives definition of knowledge elicitation as the transfer and transformation of problem solving expertise and domain knowledge from a source into a program. (McGraw. K.L., 1992). Working with the domain expert(s), the knowledge engineer elicits, encodes and continuously refines the knowledge base until an acceptable performance is not achieved.
The process of knowledge engineering involves knowledge acquisitions which includes knowledge elicitation and other activities such as knowledge explication and conceptual modeling (Regoczei, 1992), as well as the coding of the resulting knowledge the design of the usable interface, and testing and evaluating of these interfaces (Diaper, 1989b). Thus knowledge elicitation is the sub process of the knowledge engineering domain. In order to fit the knowledge elicitation into larger context of applied cognitive psychology it’s necessary to understand its brief evolution.

All knowledge is not quickly accessible; especially knowledge with symbolic description is difficult to access (Garavelli, 2002), clear distinction has to be made between explicit knowledge and the implicit knowledge. Ron Sun argued that simulating knowledge representation runs into difficulty since there is not a single type of knowledge, also mentioned that, it split into implicit and explicit knowledge, both type of knowledge represents different type of information so must be handled properly (Sun, 2003).

Research into knowledge management aims at capturing the implicit knowledge, resides into the heads of expert and making them explicit for general use. Tacit Knowledge is essentially the driving force behind such innovation as new technologies, processes or techniques (Maqsood, 2004).

With new focus on knowledge, question regarding knowledge elicitation became central to both applied and basic endeavors. How can knowledge be effectively elicited from an expert. Although the cognitive literature has addressed the issue of knowledge, the focus is on the question of its representation in to knowledge base system (Anderson, 1995).

The creation of knowledge starts with an individual and spiral through successive conversions modes: like Internalization, Socialization, Externalization and Combination. This essentially results in its amplification as large number of individuals, groups, and eventually the organization relate to the newly created knowledge (Vat K.H., 2003).

Repetition of certain tasks over time usually results in the execution of such tasks without need for conscious thought; also verbalization may not be valid description of real processes especially these processes are difficult for an expert to verbalize. Interviews also encourage experts to speculate about their cognitive processes.

There are two broad classes of knowledge elicitation manual and automation (machine learning).

3. PROBLEMS IN –ELICITATION

These are the problems encountered during proper knowledge elicitation. Good elicitation tries to identify:

The expert’s decision making strategies
The consequences of such strategies
Methods of improving on such strategies
The usually vary depending on chosen techniques for elicitation (Diaper, 1989b).
4. PROBLEMS IN AUTOMATED KNOWLEDGE ELICITATION:

Automated knowledge elicitation handles elicitation, analysis and transfer of knowledge almost simultaneously. The main emphasis is on development of knowledge base system that gives straight set of rules from past examples or samples of knowledge gained interactively from domain expert. Its saves energy, money and time for knowledge engineers.

- Interactive knowledge acquisition and encoding tools can greatly reduce the need of acting as intermediary for the knowledge engineers, but in most applications they leave a substantial number of roles for knowledge engineers as shown in Fig 1. Interactive knowledge elicitation can be combined with manual elicitation, so that with interactive available tools knowledge engineers proceeds more efficiently with experts.

- In many cases of knowledge elicitation the problem of insufficient data set arises with which the knowledge base shell is to be built, so the knowledge engineer still needs to do manual elicitation of certain first-hand expert’s knowledge from which the shell is going to be modeled. The shell contains objects, classes and pattern matching rules used to weigh and elicit further rules.

- Proper abstractions of needed facts are essential with proper skills in software engineering that would reduce redundant questions to reduce the complexity of rule set generated from it.

- Uncertainties and fake associations are encountered when experts enter incomplete or incorrect answers to some of inquiries of an automated knowledge elicitation tools. The resultant relation may appear al right but shows the illogical sequence of reasoning so the expert must be trained first about that tool. But training the expert to effectively use the automated tool may itself become bottleneck.

But in all cases, the knowledge engineer starts by getting enough background on expert’s domain so as to guide entries as well as correct errors in the knowledge base built. The expert can also learn these skills of expertise transfer by trial and error, which is supported only by well programmed elicitation environment friendly enough to guide the learning (Eric C. Okafor, 2006).
5. NEURAL NETWORK

Most of the auto-induction methods are insufficient to handle the exact correctness of certain data. Good solution to the problem is pattern matching along with weight adjusting technology called neural networks, an attempt to model neurons in human brain. It is well suited for:

- Uncertainties
- Generalization and reasoning
- Pattern matching
- Self-training and improvements and for handling
- Large and complex data sets.

6. CONCLUSION

The above discussion has exposed the different underlying problems in knowledge elicitation towards automation. In today’s functional expert systems, there is usually a strategic application of several methods to elicit, transfer and properly model the expert’s knowledge into the knowledge base system. This guarantees that it may effectively replaces the human expert in a specific domain, but knowledge elicitation problems are real, fully automation of knowledge engineering is still being pursued, given the underlying issues and difficulties in knowledge elicitation.
REFERENCES

FRAMEWORK FOR DECISION MAKING WITH TECHNOLOGICAL INFEERENCE IN MANAGEMENT

Leena Anum* and Khalil Ahmed
School of Computer Science, National College of Business Administration and Economics, Lahore, Pakistan
Email: *leena.rehman@hotmail.com

ABSTRACT

Decision making is the basic aspect in organizational management, therefore it is essential to observe the elements that are critical in order to develop strategic prospects for better performance in decision making. Technology is playing a vital role in providing support at operational and managerial levels, though the impact of technology on decision making is still under debate. There is a need to infuse technology as a basic process integrated at all managerial levels. This paper intends to provide a framework for technological inference in management structure to increase the overall efficiencies of functions related to decision making.

1. INTRODUCTION

In corporate management, decision making is always a major debate due to the factor of change which is closely dependent on rapid and accurate availability of change-relevant information. As discussed by (Murawski, 2000) in 1986, Ackerman suggested that there mainly are three types of changes, namely: developmental, transitional and transformational. Developmental being concerned with the improvement of existing aspects, whereas transitional change is considered to be one related to that of changing known aspects with prior understanding to outcomes under a specific time span. However, transformational changes are broad changes that are entirely new with that of the initiation of different and innovative capacities.

Technology in itself is a transformational change as it has brought the corporate world from raw or book listed data entries to technical and secure data clusters, statistical applications, critical analysis, process control mechanisms, online communication and global linkages. The inference of technology into management is an aspect to support and accelerate the organizational functions and improving their capacities at large. Efficiency and effectiveness, risk management and optimization, intellectual capital retention, employee performance and similar others, being some of the important aspects of organizational concerns, could be brought in a single loop by technological infusion, ultimately enabling the management to run all the processes side by side. It also adds to decision making aspect of management due to the analytical and computation characteristics. Thus, it plays a synergic role for management with the capacity of decision making in a direct manner and indirectly supports the system by improving employee performance and retention.
The role of top and middle management is critical when it comes to decision making and strategy development in an organization. This is because the complete organizational hierarchy is interlinked with each and every aspect, thus, having an impact on the other. The decisions and strategies being formulated at the top and middle level of management have significant impact on the successful running of operational functions. Setbacks at the two levels shall damage the entire loop of operational linkages and changes in any of the top two levels, creates knowledge gaps, which ultimately trickles down its effect to the operational level. In this regard, technological inference can act as a complimentary tool to the management, providing substantial support to the functions and their successful running.

In knowledge-based economy where globalization has broadened the organizational boundaries to across continents, knowledge and communication have become a necessity and considered as a serious resource for an organization. In this era of transformational changes, where technology has bypassed traditional management approaches, it is a requirement of the organizations to evaluate its technological competencies and recognize them in their annual reports.

Thus there is a dire need to equip the management levels with proficient technological competencies. The framework has been designed to provide a sketch of how and where is the infusion of technology required and how shall it benefit the management concerns at fullest.

2. LITERATURE REVIEW

Technology inference is a topic under constant debate. In the field of management where decision making and strategy formulation are critical aspects of process and control, there is a dire need to incorporate supportive technical tools and systems. Even in this age of global advancement, there still is an underutilization of technology at different levels of professional functions. It is not being utilized as per its core competencies, thus not having maximum benefit out of it. The present deficiencies in technology utilization prove to be an obstacle in the achievement of organizational objectives to the fullest. Technological infrastructure is that aspect of an organization on which various processes and sub-processes are dependent upon. (Shih, 2002)

Different organizational management processes are still unequipped with the appropriate technological inference that can not only increase the efficiency and effectiveness of the process as a whole but also ensure better performance of the employees. In the technology-enhanced era, organizations must have technological tools for business and management processes, their effective and efficient running and for the identification of expected or associated risks. Thus the infusion of technology in management shall benefit the organizational management infrastructure by adding value to it. It is the demand of the new generation management functions.(Helic, 2006).

(Helic, 2006)Various other researchers (Repman et al. 2004; Hawks, 2001; Rico, 2003; George, 2004) also highlighted the role of technological inference from operational to top management level. Organizations by now are moreover a social entity based on communication between members of different levels of management, their process achievements, analysis and decisions made for further tasks and operations.
Information technology is that comprehensive system which has lead organizations to an entirely different world of management, communication and operations. The infusion of this particular aspect to management shall increase employee performance and retention by making their jobs more innovative and practical. Transforming traditional book entry systems and communication into an entirely automated information system has opened doors to new paths for organizational success. Organizations at present are in need of this aspect in order to identify future probabilities, forecast risks and threats and communicate in between the organizational levels for strategic actions towards management and control. (Mvungi & Jay, 2009)

As discussed by (Mvungi & Jay, 2009), according to (Krippendorff, 1986), information technology is not only a tool for a single aspect of organizational management; it is in fact the basis of multiple functions being performed at various levels thus adding to growth of the organization.

Business intelligence is another utility under the helm of information technology, providing future vision to organizational management. This enables the organizations to identify the probabilities for future growth, identifying risks and threats and deducing strategic approaches to overcome the expected threats. It is such facilitation to management that provides them capacities of acquisition and analysis of data, its formulation in the form of management strategies and re-engineering functions, ultimately opening new doors to knowledge evolution and management in an organization. (Olszak & Ziemb, 2007)

It also reduces many of time constraints, external pressures, monitoring fallacies and provides a constant check on the functions, their running and associated considerations. Utilization of such technological support systems, organizations are made able enough to gain competitive advantage. (Olszak & Ziemb, 2007).

As explained by (Shaukat & Zafarullah, 2009) IT is the part and parcel of almost every aspect of professional management. Incorporation of IT systems is a must as it not only has a great impact on formulation end of an organization, such as adding to the strategic decision making but also to that of the implementation end such as data storage and analyses, monitoring and feedback functions. Their work suggests that the infusion of these technological systems to their management aspects is a matter of critical consideration for almost all the sectors of any economy. However, not only the infusion is a necessity but also the utilization of these infused systems at fullest in management. They also discussed the work of (Vasudevan, 2003; Long and Long, 1999) that within few years IT has made itself an essential of managing organizational affairs. It is more like a back bone to the management aspect of any organization.

Information technology has proved to be a boost agent for behavioral changes of management personnel such as that of their adaption to innovation and transformation, strategic functioning and interpersonal and organizational interactions. (Crutzen, Nooijer, Brouwer, Oenema, Brug, & Vries, 2008).

Therefore technological infusion in management practices shall not only be beneficial for improvement in organizational functions at various levels of management, but also a value added transformation in the aspect of employee performance, retention and
foreseeing risks to the organization. The framework is proposed to support the idea of technological infusion into management, explaining its prospects to the organizations.

3. DISCUSSION

This framework is being proposed for technological inference at managerial levels and to highlight the need of this interference as an integral part of transformation. Though the use of technology is becoming a standard practice already but the coherence of various technological techniques at various corporate structures still needs modeling and appropriate architecture to use the standard managerial practices without compromising the respective strengths and also to enhance the capability to articulate most suitable and factual decision plans.

In corporate world, knowledge has been the basis for the entire decision making process, however, the structures in which this knowledge was firstly organized, stored and analyzed for decision making has been varying with time and development. This knowledge is acquired and utilized by management concerns in two basics forms, one as a collection of past data in the form of previous knowledge and the other as a collection of present practices as current knowledge. Technological advancement specifically in the field of information technology has given an entire new direction to the decision making processes to evaluate the impact of implicit and explicit knowledge in formulation of decision options. Current business intelligence applications are using various databases/data warehouses for analytical purposes but now many business intelligence tools are gradually adapting knowledge bases to analyze situations, based on information as well as knowledge to make more realistic decision choices and analytical results. It has not only provided an immense support to decision functions but also proved beneficial and complimentary to that of the operational functions at different levels of the organizations.

Proposed framework consists of a combination of both previous knowledge based and current knowledge based expert systems as a cause, these systems varying with the specificity of functions at a certain managerial level are utilized and in a result of the facilitation provided by them, the function`s and decision making aspects are supported and enhanced. As all these levels are interlinked from both top to bottom and bottom to top management interactions, the increased efficiency and functionality of one level adds to that of the proceeding levels.

Operational management

The operational level can utilize OR (Operations research) and DWH (Data warehousing) approaches as a source of previous knowledge and OR (Operations research) and ERP (Enterprise Resource Planning) as that of the current knowledge support systems. In result of utilizing these technological inferences, the operational management can improve the performance of its employees in various operations being carried out at this level of management. Improving employee performance shall ultimately add on to that of operational accomplishment, thus making the overall functions of this levels run smooth and with maximum precision.
Middle management

Technological support systems such as ANN (Artificial neural networks), D. Mining (Data mining), OR (Operations research), DWH (Data warehousing) and BI (Business intelligence) if utilized as a source of previous knowledge by the middle managers of an organization and that of BI (Business intelligence), ODB (Online data bases) and ERP (Enterprise resource planning) as current knowledge support systems, can enhance the working capacities of the intellectual capital at this level. It can benefit the middle managers by providing them better foresight for command and control, making data generation and analysis easier and reliable, thus increasing the overall efficiency of the functions being carried out at this level.

Top management

Utilizing ERP (Enterprise resource planning) as a previous knowledge support system and EXP. SYSTEM (Expert system) and BI (Business intelligence) as that of current knowledge can benefit the top management by adding on to their decision making expertise and identifying the risks associated to those decisions. These technological inferences in management shall enhance the capacities of top level managers by improving risk management, decision making, forecasting and effectiveness.

CONCLUSION

Efficiency and effectiveness are one of the foremost aspects of organizational concerns at present. Both these aspects are achievable yet demanding and challenging. To meet the challenges of customer demands along with the maintenance of organizational concerns such as performance and profitability, technological inference is a must. This is because at different levels of organizational management, from data generation to analysis, from command and control to performance management, from decision making to strategy development, every function requires to be well equipped with time saving, efficient and reliable system of practice. These systems are now available in the form of technological support systems such as those defined in the framework proposed for technological inference in management.

By incorporating technology in day to day management of an organization, it can not only achieve better in its working capacities but also achieve a competitive advantage on others with lesser or no use of technological support systems. It can also add to the global economy as it is beneficial in reducing time taking procedures into technologically equipped efficient systems.

Therefore inference of technology in management is a critical requirement of organizations at present. Utilizing technological support systems shall benefit the economy at macro and individual organizations at micro level, thus creating a technologically advanced culture for management practices in organizations.
REFERENCES


AN IMPROVED CLASS OF REGRESSION ESTIMATORS
FOR THE POPULATION MEAN UNDER MULTI-PHASE SAMPLING
SCHEME IN THE PRESENCE OF NON-RESPONSE

Saba Riaz and Javaid Shabbir
Department of Statistics, Quaid-e-Azam University, Islamabad, Pakistan.
Email: sabaqau@gmail.com; js_qau@yahoo.com

ABSTRACT

In this paper we propose a class of regression estimators to estimate the population mean $\bar{Y}$ of the study variable $y$ using multi auxiliary information under multi-phase sampling scheme whenever the problem of non-response exists. The bias and mean square error of the estimators are obtained and the properties are studied for the optimum class of the proposed estimators. We also illustrate the efficiency comparison of the proposed class with the usual existing regression type estimators. Mathematical study is carried out to support the proposed class of estimators.

INTRODUCTION:

The issue of non-response in the sample surveys is very common especially in mail surveys rather than in personal interview surveys. Hansen and Hurwitz (1946) considered in the problem of non-response and proposed a technique for adjusting for non-response to address the bias problem. In estimating the population parameters, use of auxiliary information improves the precision of the estimators. Furthermore, Okafer and Lee (2000), Tabasum and Khan (2004, 2006), Singh and Kumar (2008, 2009, 2010, 2011) have suggested many improved estimation techniques in the problems of non-response using the information on auxiliary variables. Diana et al. (2004, 2007) have proposed a general class of estimators of population using auxiliary information under multi phase sampling scheme.

By using Hansen and Hurwitz technique:

Consider a finite population of size $N$ and a random sample of size $n$ is drawn by using without replacement sampling scheme. While conducting surveys for population parameter estimation, $n_1$ sample units response to the items, but the remaining $(n-n_1)$ units do not provide required information.

First, select a sample of size $n$, a questionnaire is mailed to the sampling units.

A sub sample of size $r = \frac{n_2}{k}$, where $k \geq 1$ from the $n_2$ non-responding units in the initial sampling is contacted through personal interviews.
Consider the population of \( N \) into two groups or strata, named as response and non-response. The size of the stratum response is \( N_1 \) and \( N_2 \) is the size of the stratum “non-response” in the population, where \( N_2 = N - N_1 \).

Let, \( (y_1, y_2, \ldots, y_{N_1}) \) units are from the respondents group and \( (y_{N_1+1}, \ldots, y_{N_1+N_2}) \) units are from the non-respondents group.

The population mean is \( \bar{Y} = \frac{\sum_{i=1}^{N} y_i}{N} \),

\[ \bar{Y}_1 = \frac{\sum_{i=1}^{N_1} y_i}{N_1} \text{ is mean of response group and } \bar{Y}_2 = \frac{\sum_{i=N_1+1}^{N} y_i}{N_2} \text{ is mean of non-response group.} \]

An unbiased estimator for the population mean \( \bar{Y} \) is \( \bar{y} = w_1 \bar{Y}_1 + w_2 \bar{Y}_2 \)

where, \( \bar{y}_1 = \frac{n_1}{n} \bar{Y}_1 \), \( \bar{y}_2 = \frac{r}{n} \bar{Y}_2 \), \( w_1 = \frac{n_1}{n} \) and \( w_2 = \frac{n_2}{n} \). The variance of \( \bar{y} \) is given by

\[ Var(\bar{y}) = \left( \frac{1-f}{n} \right) S_y^2 + \frac{W_2(k-1)}{n} S_{y(2)}^2, \text{ where, } f = \frac{n}{N}, \ S_y^2 = \frac{\sum_{i=1}^{N} (y_i - \bar{Y})^2}{N-1} \]

and \( S_{y(2)}^2 = \frac{\sum_{i=N_1+1}^{N} (y_i - \bar{Y}_2)^2}{N_2-1} \)

Similarly, the unbiased estimator of the population mean \( \bar{X} \) of the auxiliary variate \( x \) is \( \bar{x} = w_1 \bar{x}_1 + w_2 \bar{x}_2 \), where \( \bar{x}_1 = \frac{n_1}{r} \bar{X}_1 \) and \( \bar{x}_2 = \frac{r}{r} \bar{X}_2 \)

The variance of \( \bar{x} \) is given by \( Var(\bar{x}) = \left( \frac{1-f}{n} \right) S_x^2 + \frac{W_2(k-1)}{n} S_{x(2)}^2 \), where,

\[ S_x^2 = \frac{\sum_{i=1}^{N} (x_i - \bar{X})^2}{N-1}, \text{ and } S_{x(2)}^2 = \frac{\sum_{i=N_1+1}^{N} (x_i - \bar{X}_2)^2}{N_2-1}. \]

CLASS OF PROPOSED ESTIMATORS AND THEIR MEAN SQUARE ERROR:

We are proposing here a class of regression estimators by following procedures.
To estimate the population mean of \( y \) under multi-phase sampling, let \( n_1 (n_1 < N) \) is the first sample drawn from population by a simple random sampling without replacement (SRSWOR), \( n_2 (n_2 < n_1) \) is the second sub sample drawn from \( N_1 \) non-respondents group by SRSWOR, \( n_3 (n_3 < n_2 < n_1) \) is the third sub sample drawn from \( N_2 \) non-respondents group by SRSWOR as well and so on up to \( n_{k+1} (n_{k+1} < n_k < \cdots < n_1) \) the last sample sub sample is drawn from \( N_k \) non-respondents group by SRSWOR which is the smallest one. Here, we have \( k \) auxiliary variables and \((k+1)\) phases.

**Situation 1:**

In first situation, each auxiliary variable is observed only once at the phase where is recorded at the first time. Assume \( \bar{X}_1, \bar{X}_2, \ldots, \bar{X}_k \) are known auxiliary variables, also non-response exist in both study and auxiliary variables.

\[
\hat{T}_{11} = \bar{Y}_{0(k+1)} + \nu_1 \left( \bar{X}_1 - \bar{x}_l(1) \right) + \nu_2 \left( \bar{X}_2 - \bar{x}_l(2) \right) + \nu_3 \left( \bar{X}_3 - \bar{x}_l(3) \right) + \cdots + \nu_{k-1} \left( \bar{X}_{k-1} - \bar{x}_{k-l(k-1)} \right) + \nu_k \left( \bar{X}_k - \bar{x}_l(k) \right)
\]

where, \( \nu_1, \nu_2, \ldots, \nu_k \) are constants.

To find the bias and MSE of the estimator \( \hat{T}_{11} \) up to first order of approximation, we expand the above equation in terms of \( e' s \),

\[
\hat{T}_{11} = \bar{Y} + \bar{Y}e_{0(k+1)} - \nu_1 \bar{X}_1 e_{l(1)} - \nu_2 \bar{X}_2 e_{l(2)} - \nu_3 \bar{X}_3 e_{l(3)} - \cdots - \nu_{k-1} \bar{X}_{k-1} e_{l(k-1)} + \nu_k \bar{X}_k e_{l(k)}
\]

Taking expectation of (2.2), we obtain

\[
B(\hat{T}_{11}) = E(\hat{T}_{11} - \bar{Y}) = 0
\]

(2.3)

The mean square error of the proposed estimator \( \hat{T}_{11} \), up to first order of approximation

\[
MSE(\hat{T}_{11}) = E\left[ \bar{Y}e_{0(k+1)} - \nu_1 \bar{X}_1 e_{l(1)} - \nu_2 \bar{X}_2 e_{l(2)} - \nu_3 \bar{X}_3 e_{l(3)} - \cdots - \nu_{k-1} \bar{X}_{k-1} e_{l(k-1)} + \nu_k \bar{X}_k e_{l(k)} \right]^2
\]

(2.4)
\[ MSE(\hat{T}_{11}) = \left( f_{k+1}S_{y}^2 + w_{k+1}S_{y(2)}^2 \right) + v_1^2 \left( f_{1}S_{x_1}^2 + w_{1}S_{x_1(2)}^2 \right) + v_2^2 \left( f_{2}S_{x_2}^2 + w_{2}S_{x_2(2)}^2 \right) + v_3^2 \left( f_{3}S_{x_3}^2 + w_{3}S_{x_3(2)}^2 \right) + \cdots + v_{k-1}^2 \left( f_{k-1}S_{x_{k-1}}^2 + w_{k-1}S_{x_{k-1}(2)}^2 \right) + v_k^2 \left( f_{k}S_{x_k}^2 + w_{k}S_{x_k(2)}^2 \right) - 2v_1 \left( f_{1}S_{y_{x_1}} + w_{1}S_{y_{x_1}(2)} \right) - 2v_2 \left( f_{2}S_{y_{x_2}} + w_{2}S_{y_{x_2}(2)} \right) - \cdots - 2v_{k-1} \left( f_{k-1}S_{y_{x_{k-1}}} + w_{k-1}S_{y_{x_{k-1}(2)}} \right) - 2v_k \left( f_{k}S_{y_{x_k}} + w_{k}S_{y_{x_k}(2)} \right) + 2v_1v_2 \left( f_{1}S_{x_{1}x_2} + w_{1}S_{x_{1}x_2(2)} \right) + 2v_1v_3 \left( f_{1}S_{x_{1}x_3} + w_{1}S_{x_{1}x_3(2)} \right) + \cdots + 2v_1v_k \left( f_{1}S_{x_{1}x_k} + w_{1}S_{x_{1}x_k(2)} \right) + 2v_2v_3 \left( f_{2}S_{x_{2}x_3} + w_{2}S_{x_{2}x_3(2)} \right) + 2v_2v_4 \left( f_{2}S_{x_{2}x_4} + w_{2}S_{x_{2}x_4(2)} \right) + \cdots + 2v_2v_k \left( f_{2}S_{x_{2}x_k} + w_{2}S_{x_{2}x_k(2)} \right) + \cdots + 2v_{k-1}v_k \left( f_{k-1}S_{x_{k-1}x_k} + w_{k-1}S_{x_{k-1}x_k(2)} \right) \]

where, \( f_i = \left( \frac{1}{n_i} - \frac{1}{N_i} \right) \), \( w_j = \frac{W_i(k-1)}{n_i} \), \( W_i = \frac{N_i}{N} \), \( i = 1, 2, \ldots, k + 1 \)

Simplifying above equation, we can write it as

\[ MSE(\hat{T}_{11}) = S_{y}^2 + v_1^2S_{x_1}^2 + v_2^2S_{x_2}^2 + v_3^2S_{x_3}^2 + \cdots + v_{k-1}^2S_{x_{k-1}}^2 + v_k^2S_{x_k}^2 - 2v_1S_{y_{x_1}} - 2v_2S_{y_{x_2}} - \cdots - 2v_{k-1}S_{y_{x_{k-1}}} - 2v_kS_{y_{x_k}} + 2v_1v_2S_{x_{1}x_2} + 2v_1v_3S_{x_{1}x_3} + \cdots + 2v_1v_kS_{x_{1}x_k} + 2v_2v_3S_{x_{2}x_3} + \cdots + 2v_2v_kS_{x_{2}x_k} + \cdots + 2v_{k-1}v_kS_{x_{k-1}x_k} \]

Differentiating (2.6) w.r.t \( v_i \)'s and equating to zero, where \( i = 1, 2, \ldots, k \)

\[
\begin{pmatrix}
S_{x_1}^2 & S_{x_1x_2}^* & S_{x_1x_3}^* & \ldots & S_{x_1x_{k-1}}^* & S_{x_1x_k}^* \\
S_{x_1x_2}^* & S_{x_2x_2}^* & S_{x_2x_3}^* & \ldots & S_{x_2x_{k-1}}^* & S_{x_2x_k}^* \\
S_{x_1x_3}^* & S_{x_2x_3}^* & S_{x_3x_3}^* & \ldots & S_{x_3x_{k-1}}^* & S_{x_3x_k}^* \\
\vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\
S_{x_1x_{k-1}}^* & S_{x_2x_{k-1}}^* & S_{x_3x_{k-1}}^* & \ldots & S_{x_{k-1}x_{k-1}}^* & S_{x_{k-1}x_k}^* \\
S_{x_1x_k}^* & S_{x_2x_k}^* & S_{x_3x_k}^* & \ldots & S_{x_{k-1}x_k}^* & S_{x_k}^2
\end{pmatrix}
\begin{pmatrix}
v_1 \\
v_2 \\
v_3 \\
\vdots \\
v_{k-1} \\
v_k
\end{pmatrix}
= \begin{pmatrix}
S_{y_{x_1}}^* \\
S_{y_{x_2}}^* \\
S_{y_{x_3}}^* \\
\vdots \\
S_{y_{x_{k-1}}}^* \\
S_{y_{x_k}}^* 
\end{pmatrix}
\]

defining above matrix as:
\[
\Sigma_{k\times k} = \begin{pmatrix}
S_{x_1}^* & S_{x_1 x_2}^* & S_{x_1 x_3}^* & \cdots & S_{x_1 x_{k-1}}^* & S_{x_1 x_k}^* \\
S_{x_2 x_1}^* & S_{x_2}^* & S_{x_2 x_3}^* & \cdots & S_{x_2 x_{k-1}}^* & S_{x_2 x_k}^* \\
S_{x_3 x_1}^* & S_{x_3 x_2}^* & S_{x_3}^* & \cdots & S_{x_3 x_{k-1}}^* & S_{x_3 x_k}^* \\
\vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\
S_{x_{k-1} x_1}^* & S_{x_{k-1} x_2}^* & S_{x_{k-1} x_3}^* & \cdots & S_{x_{k-1} x_{k-1}}^* & S_{x_{k-1} x_k}^* \\
S_{x_k x_1}^* & S_{x_k x_2}^* & S_{x_k x_3}^* & \cdots & S_{x_k x_{k-1}}^* & S_{x_k}^* 
\end{pmatrix},
\]

\[
A_{k\times 1} = \begin{pmatrix}
u_1 \\
u_2 \\
u_3 \\
\vdots \\
u_{k-1} \\
u_k
\end{pmatrix}, \quad P_{k\times 1} = \begin{pmatrix}
S_{y_1 x_1}^* \\
S_{y_1 x_2}^* \\
S_{y_1 x_3}^* \\
\vdots \\
S_{y_1 x_{k-1}}^* \\
S_{y_1 x_k}^*
\end{pmatrix},
\]

\[
A_{k\times 1} = \Sigma_{k\times k}^{-1} P_{k\times 1}
\]

\[
A_{k\times 1} = \frac{\text{Adj}(\Sigma_{k\times k}) P_{k\times 1}}{\Sigma_{1\times k}}
\]

\[
\text{Adj}(\Sigma_{k\times k}) P_{k\times 1} = \begin{bmatrix}
(-1)^{1+1} \left| \Sigma_{y_1 x_1} \right|_{y_2 x_k} \\
(-1)^{2+1} \left| \Sigma_{y_1 x_2} \right|_{y_2 x_k} \\
\vdots \\
(-1)^{k+1} \left| \Sigma_{y_1 x_k} \right|_{y_2 x_k}
\end{bmatrix}
\]

\[
\Sigma_{y_{x_1}} = \begin{vmatrix}
S_{y_1 x_1}^* & S_{y_1 x_2}^* & S_{y_1 x_3}^* & \cdots & S_{y_1 x_{k-1}}^* & S_{y_1 x_k}^* \\
S_{y_2 x_1}^* & S_{y_2}^* & S_{y_2 x_3}^* & \cdots & S_{y_2 x_{k-1}}^* & S_{y_2 x_k}^* \\
S_{y_3 x_1}^* & S_{y_3 x_2}^* & S_{y_3}^* & \cdots & S_{y_3 x_{k-1}}^* & S_{y_3 x_k}^* \\
\vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\
S_{y_{k-1} x_1}^* & S_{y_{k-1} x_2}^* & S_{y_{k-1} x_3}^* & \cdots & S_{y_{k-1} x_{k-1}}^* & S_{y_{k-1} x_k}^* \\
S_{y_k x_1}^* & S_{y_k x_2}^* & S_{y_k x_3}^* & \cdots & S_{y_k x_{k-1}}^* & S_{y_k}^*
\end{vmatrix}
\]
An improved class of regression estimators for the population

\[ \Sigma_{y_{x_2}} | y_{x_2} \]

\[
\begin{bmatrix}
S_{y_1} & S_{x_1}^{*} & S_{x_1 x_2}^{*} & \ldots & S_{x_1 x_{k-1}}^{*} & S_{x_1 x_k}^{*} \\
S_{y_2}^{*} & S_{x_1 x_2}^{*} & S_{x_2 x_3}^{*} & \ldots & S_{x_2 x_{k-1}}^{*} & S_{x_2 x_k}^{*} \\
\vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\
S_{y_{x_{k-1}}}^{*} & S_{x_1 x_{k-1}}^{*} & S_{x_2 x_{k-1}}^{*} & \ldots & S_{x_{k-2} x_{k-1}}^{*} & S_{x_{k-2} x_k}^{*} \\
S_{y_k}^{*} & S_{x_1 x_k}^{*} & S_{x_2 x_k}^{*} & \ldots & S_{x_{k-2} x_k}^{*} & S_{x_{k-1} x_k}^{*}
\end{bmatrix}
\]

And so on up to

\[ \Sigma_{y_{x_k}} | y_{x_k} \]

\[
\begin{bmatrix}
S_{y_1}^{*} & S_{x_1}^{*} & S_{x_1 x_2}^{*} & \ldots & S_{x_1 x_{k-2}}^{*} & S_{x_1 x_{k-1}}^{*} \\
S_{y_2}^{*} & S_{x_1 x_2}^{*} & S_{x_2 x_3}^{*} & \ldots & S_{x_2 x_{k-2}}^{*} & S_{x_2 x_{k-1}}^{*} \\
\vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\
S_{y_{x_{k-1}}}^{*} & S_{x_1 x_{k-1}}^{*} & S_{x_2 x_{k-1}}^{*} & \ldots & S_{x_{k-2} x_{k-1}}^{*} & S_{x_{k-2} x_k}^{*} \\
S_{y_k}^{*} & S_{x_1 x_k}^{*} & S_{x_2 x_k}^{*} & \ldots & S_{x_{k-2} x_k}^{*} & S_{x_{k-1} x_k}^{*}
\end{bmatrix}
\]

After simplifying above matrix, we get min. value of \( MSE(\hat{T}_{11}) \)

\[
\begin{bmatrix}
\nu_1 \\
\nu_2 \\
\nu_3 \\
\vdots \\
\nu_{k-1} \\
\nu_k
\end{bmatrix} = \frac{1}{\left| \Sigma \right|_{y_{x_k}}} \begin{bmatrix}
(-1)^{i+1} \left| \Sigma_{y_{x_i}} | y_{x_k} \right| \\
(-1)^{i+1} \left| \Sigma_{y_{x_i}} | y_{x_k} \right| \\
\vdots \\
(-1)^{i+1} \left| \Sigma_{y_{x_i}} | y_{x_k} \right|
\end{bmatrix}
\]

\[
\nu_i = \frac{(-1)^{i+1} \left| \Sigma_{y_{x_i}} | y_{x_k} \right|}{\left| \Sigma \right|_{y_{x_k}}}, \quad i = 1, 2, \ldots, k
\]

We can write eq. (2.6) as:

\[
MSE(\hat{T}_{11}) = S_{y_1}^{*2} + \sum_{i=1}^{k} \nu_i^2 S_{x_i}^{*2} - 2 \sum_{i=1}^{k} \nu_i S_{y_{x_i}}^{*} + 2 \sum_{i=1}^{k-1} \sum_{j>i}^{k} \nu_i \nu_j S_{x_{i,x_j}}^{*}
\]
\[
MSE(\hat{T}_{11})_{\min} = S_y^* + \sum_{i=1}^{k} \left( \frac{(-1)^{i+1} |\Sigma_{yx_i}| |y_{x_k}|}{|\Sigma|_{x_k}} \right)^2 S_{x_i}^* - 2 \sum_{i=1}^{k} \left( \frac{(-1)^{i+1} |\Sigma_{yx_i}| |y_{x_k}|}{|\Sigma|_{x_k}} \right) S_{x_i}^* \\
+ 2 \sum_{i=1}^{k-1} \sum_{j>i} \left( \frac{(-1)^{i+1} |\Sigma_{yx_i}| |y_{x_k}|}{|\Sigma|_{x_k}} \right) \left( \frac{(-1)^{j+1} |\Sigma_{yx_j}| |y_{x_k}|}{|\Sigma|_{x_k}} \right) S_{x_i,x_j}^*
\]

\[
MSE(\hat{T}_{11})_{\min} = S_y^2 + \frac{1}{\left( |\Sigma|_{x_k} \right)^2} \left[ \sum_{i=1}^{k} \left( \frac{(-1)^{i+1} |\Sigma_{yx_i}| |y_{x_k}|}{|\Sigma|_{x_k}} \right)^2 S_{x_i}^* - 2 \sum_{i=1}^{k} \left( \frac{(-1)^{i+1} |\Sigma_{yx_i}| |y_{x_k}|}{|\Sigma|_{x_k}} \right) S_{x_i}^* \\
+ \sum_{i=1}^{k-1} \sum_{j>i} \left( \frac{(-1)^{i+1} |\Sigma_{yx_i}| |y_{x_k}|}{|\Sigma|_{x_k}} \right) \left( \frac{(-1)^{j+1} |\Sigma_{yx_j}| |y_{x_k}|}{|\Sigma|_{x_k}} \right) S_{x_i,x_j}^* \right] S_y^2 \left( \frac{1}{\left| \Sigma \right|_{x_k}^2} \right)^2 S_{x_i}^* \left( \frac{1}{\left| \Sigma \right|_{x_k}^2} \right)^2 S_{x_i,x_j}^*
\]

\[
MSE(\hat{T}_{11})_{\min} = S_y^2 \left( 1 - R_{y,x_k}^2 \right) \quad (2.8)
\]

A Partial Use of Auxiliary Information:

If we take two auxiliary variables \( x_1 \) and \( x_2 \) (both have non-response) using three phase sampling scheme, at first phase we estimate \( \bar{x}_1^* \), at second phase we estimate \( \bar{x}_2^* \), and at third phase we estimate \( \bar{y}^* \), as given by

\[
\hat{T}_{11(2)} = \bar{y}^* + u_1 \left( \bar{x}_1 - \bar{x}_{1(1)} \right) + u_2 \left( \bar{x}_2 - \bar{x}_{2(2)} \right) \quad (2.9)
\]

\[
MSE(\hat{T}_{11(2)}) = E \left[ \left( \bar{y}^* - u_1 \bar{x}_{1(1)}^* - u_2 \bar{x}_{2(2)}^* \right)^2 \right] \quad (2.10)
\]

\[
MSE(\hat{T}_{11(2)}) = \left( f_3 S_y^2 + w_3 S_{y(2)}^2 \right) + u_1^2 \left( f_1 S_{x_1}^2 + w_1 S_{x_1(2)}^2 \right) \\
+ u_2^2 \left( f_2 S_{x_2}^2 + w_2 S_{x_2(2)}^2 \right) - 2 u_1 \left( f_1 S_{y,x_1} + w_1 S_{y,x_1(2)} \right) \\
- 2 u_2 \left( f_2 S_{y,x_2} + w_2 S_{y,x_2(2)} \right) + 2 u_1 u_2 \left( f_1 S_{x_1,x_2} + w_1 S_{x_1,x_2(2)} \right) \quad (2.11)
\]
\[ \text{MSE}(\hat{T}_{1l(2)}) = S_y^{*2} + \nu_1 S_{x1}^{*2} + \nu_2 S_{x2}^{*2} - 2\nu_1 S_{y1}^{*} - 2\nu_2 S_{y2}^{*} + 2\nu_1 \nu_2 S_{x1x2}^{*} \quad (2.12) \]

where,
\[
S_y^{*2} = \left( f_3 S_y^{2} + w_3 S_y^{2}(2) \right), \quad S_{x1}^{*2} = f_1 S_{x1}^{2} + w_1 S_{x1}(2), \quad S_{x2}^{*2} = f_2 S_{x2}^{2} + w_2 S_{x2}(2), \\
S_{y1}^{*} = f_1 S_{y1} + w_1 S_{y1}(2), \quad S_{y2}^{*} = f_2 S_{y2} + w_2 S_{y2}(2), \quad S_{x1x2}^{*} = f_1 S_{x1x2} + w_1 S_{x1x2}(2), \\
f_i = \left( \frac{1}{n_i} \right), \quad w_j = \frac{W_i (k-1)}{n_i}, \quad W_i = \frac{N_i}{N}, \quad i = 1, 2, 3.
\]

Differentiating (2.12) w.r.t \( \nu_i \)'s and equating to zero, where \( i = 1, 2 \)
\[
\nu_1 = \frac{S_{x1}^{*2} S_{x2}^{*} - S_{y1}^{*2} S_{x1x2}^{*}}{S_{x1}^{2} S_{x2}^{2} - S_{x1x2}^{2}} = \nu_1(0) \text{ (say)}, \quad \nu_2 = \frac{S_{x1}^{*2} S_{x2}^{*} - S_{y1}^{*2} S_{x1x2}^{*}}{S_{x1}^{2} S_{x2}^{2} - S_{x1x2}^{2}} = \nu_2(0) \text{ (say)}
\]

\[ \text{MSE}(\hat{T}_{1l(2)})_{\text{min}} = S_y^{*2} - \left( \frac{S_{x1}^{*2} S_{x2}^{*} + S_{y1}^{*2} S_{y2}^{*} - 2S_{y1}^{*} S_{y2}^{*} S_{x1x2}^{*}}{S_{x1}^{*2} S_{x2}^{*} - S_{x1x2}^{*2}} \right) \quad (2.13) \]

\[ \text{MSE}(\hat{T}_{1l(2)})_{\text{min}} = S_y^{*2} \left( 1 - \frac{\rho_{y1}^{*2} + \rho_{y2}^{*2} - 2\rho_{y1}^{*} \rho_{y2}^{*} \rho_{y1x2}^{*}}{1 - \rho_{y1x2}^{*2}} \right) \quad (2.14) \]

\[ \text{MSE}(\hat{T}_{1l(2)})_{\text{min}} = S_y^{*2} \left( 1 - R_{y1x2}^{*2} \right) \quad (2.15) \]

**Situation 2:**

In this case, each auxiliary variable is observed twice, first it is observed at the phase where is recorded at the first time with full response and secondly its at just following phase with non-response

\[
\hat{t}_{k(1)} = \bar{y} + \alpha_1 (\bar{x}_{l(1)} - \bar{x}_{l(2)}) + \alpha_2 (\bar{x}_{2(2)} - \bar{x}_{2(3)}) + \alpha_3 (\bar{x}_{3(3)} - \bar{x}_{3(4)}) \\
+ \ldots + \alpha_{k-1} (\bar{x}_{k-1(l-1)} - \bar{x}_{k-1(l)}) + \alpha_k (\bar{x}_{k(k)} - \bar{x}_{k(k+1)}) \quad (3.1)
\]

where, \( \alpha_1, \alpha_2, \ldots, \alpha_k \) are constants. To find the bias and MSE of the estimator \( \hat{t}_{k(1)} \) up to first order of approximation, we expand the above equation in terms of \( e^* \)'s

\[
\hat{t}_{k(1)} = \bar{Y} + \bar{Y} e_{0(k+1)} + \alpha_1 \bar{X}_1 (e_{l(1)} - e_{l(2)}) + \alpha_2 \bar{X}_2 \left( e_{2(2)} - e_{2(3)} \right) + \alpha_3 \bar{X}_3 \left( e_{3(3)} - e_{3(4)} \right) \\
+ \ldots + \alpha_{k-1} \bar{X}_{k-1} (e_{k-1(l-1)} - e_{k-1(l)}) + \alpha_k \bar{X}_k \left( e_{k(k)} - e_{k(k+1)} \right) \quad (3.2)
\]

\[ B(\hat{t}_{k(1)}) = E(\hat{t}_{k(1)} - \bar{Y}) = 0 \quad (3.3) \]
The mean square error of the proposed estimator $\hat{f}_{k(11)}$, up to first order of approximation

$$MSE(\hat{f}_{k(11)}) = E\left[ \frac{1}{N} e_{0(k+1)}^2 + \alpha_1^2 \bar{x}_1^2 \left( e_{i(1)} - e_{i(2)}^* \right)^2 + \alpha_3^2 \bar{x}_2^2 \left( e_{i(3)} - e_{i(4)}^* \right)^2 + \alpha_k^2 \bar{x}_k^2 \left( e_{i(k)} - e_{i(k+1)}^* \right)^2 \right]$$

$$+ \alpha_3^2 \bar{x}_3^2 \left( e_{s(3)} - e_{s(4)}^* \right)^2 + \ldots + \alpha_{k-1}^2 \bar{x}_{k-1}^2 \left( e_{k-1(k-1)} - e_{k-1(k)}^* \right)^2 + \alpha_k^2 \bar{x}_k^2 \left( e_{k(k)} - e_{k(k+1)}^* \right)^2$$

$$+ 2 \bar{x}_1 \left( e_{i(1)} - e_{i(2)}^* \right) \left( e_{i(2)} - e_{i(2)}^* \right) + \ldots + \alpha_k \bar{x}_k \left( e_{k(k)} - e_{k(k+1)}^* \right)$$

$$MSE(\hat{f}_{k(11)}) = f_{k+1} S_y^2 + w_{k+1} S_{y(2)}^2 + \alpha_1^2 \left( (f_2 - f_1) S_{x_1}^2 + w_2 S_{x_1(2)}^2 \right)$$

$$+ \alpha_1^2 \left( (f_3 - f_2) S_{x_2}^2 + w_3 S_{x_2(2)}^2 \right) + \alpha_3^2 \left( (f_4 - f_3) S_{x_3}^2 + w_4 S_{x_3(2)}^2 \right)$$

$$+ \ldots + \alpha_2^2 \left( (f_{k+1} - f_k) S_{x_{k-1}}^2 + w_k S_{x_{k-1}(2)}^2 \right)$$

$$+ \alpha_k^2 \left( (f_{k+1} - f_k) S_{x_k}^2 + w_{k+1} S_{x_k(2)}^2 \right) - 2 \alpha_1 \left( (f_2 - f_1) S_{y_{x_1}} + w_2 S_{y_{x_1}(2)} \right)$$

$$- 2 \alpha_2 \left( (f_3 - f_2) S_{y_{x_2}} + w_3 S_{y_{x_2}(2)} \right) - 2 \alpha_3 \left( (f_4 - f_3) S_{y_{x_3}} + w_4 S_{y_{x_3}(2)} \right)$$

$$- \ldots - 2 \alpha_k \left( (f_{k+1} - f_k) S_{y_{x_{k-1}}} + w_{k+1} S_{y_{x_{k-1}}(2)} \right)$$

$$- 2 \alpha_k \left( (f_{k+1} - f_k) S_{y_{x_k}} + w_{k+1} S_{y_{x_k}(2)} \right)$$

$$MSE(\hat{f}_{k(11)}) = S_y^2 + \alpha_1 S_{x_1}^2 + \alpha_2 S_{x_2}^2 + \alpha_3 S_{x_3}^2 + \ldots + \alpha_{k-1} S_{x_{k-1}}^2 + \alpha_k S_{x_k}^2$$

$$- 2 \alpha_1 S_{y_{x_1}} - 2 \alpha_2 S_{y_{x_2}} - 2 \alpha_3 S_{y_{x_3}} - \ldots - 2 \alpha_{k-1} S_{y_{x_{k-1}}} - 2 \alpha_k S_{y_{x_k}}$$

$$MSE(\hat{f}_{k(11)}) = S_y^2 + \sum_{i=1}^{k} \alpha_i S_{x_i}^2 - 2 \sum_{i=1}^{k} \alpha_i S_{y_{x_i}}$$

where, $f_i = \left( \frac{1}{n_i} - 1 \right), \quad w_i = \frac{W_i (k-1)}{n_i}, \quad W_i = \frac{N_i}{N}, \quad i = 1, 2, \ldots, k+1$

Simplifying the above equation, we can write (3.5) as,

$$MSE(\hat{f}_{k(11)}) = S_y^2 + \sum_{i=1}^{k} \alpha_i S_{x_i}^2 - 2 \sum_{i=1}^{k} \alpha_i S_{y_{x_i}}$$

Differentiating (3.7) w.r.t $\alpha_i$’s and equating to zero, where $i = 1, 2, \ldots, k$. 

Saba and Shabbir
\[ \alpha_i = \frac{S_{y_i}^*}{S_{x_i}^2} = \alpha_{0i} \text{ (say)} \]

\[ \text{MSE} \left( \hat{t}_{k(1)} \right)_{\text{opt}} = S_{y}^* \left( 1 - \frac{k}{i=1} \rho_{y_i}^2 \right), \quad (3.8) \]

where, \( \rho_{y_i}^2 = \frac{S_{y_i}^2}{S_{y}^* S_{x_i}^*} \), \( i = 1, 2, \ldots, k \)

**A Partial Use of Auxiliary Information:**

If we take two auxiliary variables \( x_1 \) and \( x_2 \) (both have non-response) using three phase sampling scheme, we estimate \( \bar{x}_1 \) and \( \bar{x}_1^* \) is estimated at second phase, then at second phase we estimate \( \bar{x}_2 \) and \( \bar{x}_2^* \) is estimated at third phase, and at third phase we estimate \( \bar{y}^* \), as given by

\[ \hat{t}_{2(1)} = \bar{y}_1^* + \alpha_1 \left( \bar{x}_1(1) - \bar{x}_1^*(2) \right) + \alpha_2 \left( \bar{x}_2(2) - \bar{x}_2^*(3) \right) \quad (3.9) \]

\[ B \left( \hat{t}_{2(1)} \right) = E \left( \hat{t}_{2(1)} - \bar{Y} \right) = 0 \quad (3.10) \]

\[ \text{MSE} \left( \hat{t}_{2(1)} \right) = E \left( \hat{t}_{2(1)} - \bar{Y} \right)^2 = E \left[ \bar{Y}_{e(3)}^* + \alpha_1 \bar{x}_1 \left( e_{i(1)} - e_{i(2)}^* \right) + \alpha_2 \bar{x}_2 \left( e_{i(2)} - e_{i(3)}^* \right) \right]^2 \]

\[ \text{MSE} \left( \hat{t}_{2(1)} \right) = f_3 S_y^2 + w_3 S_{y(2)}^2 + \alpha_1 \left( (f_2 - f_1) S_{x_1}^2 + w_2 S_{x_1(2)}^2 \right) \]

\[ + \alpha_2 \left( (f_3 - f_2) S_{x_2}^2 + w_3 S_{x_2(2)}^2 \right) - 2 \alpha_1 \left( (f_2 - f_1) S_{y_1} + w_2 S_{y_1(2)} \right) \]

\[ - 2 \alpha_2 \left( (f_3 - f_2) S_{y_2} + w_3 S_{y_2(2)} \right) \]

\[ \quad (3.11) \]

We can write it in simplified form,

\[ \text{MSE} \left( \hat{t}_{2(1)} \right) = S_{y}^* + \alpha_1 S_{x_1}^* + \alpha_2 S_{x_2}^* - 2 \alpha_1 S_{y_1} + 2 \alpha_2 S_{y_2} \]

\[ \quad (3.12) \]

where,

\[ S_{y}^* = f_3 S_y^2 + w_3 S_{y(2)}^2, \quad S_{x_1}^* = (f_2 - f_1) S_{x_1}^2 + w_2 S_{x_1(2)}^2, \]

\[ S_{x_2}^* = (f_3 - f_2) S_{x_2}^2 + w_3 S_{x_2(2)}^2, \quad S_{y_1}^* = (f_2 - f_1) S_{y_1} + w_2 S_{y_1(2)} \]

\[ S_{y_2}^* = (f_3 - f_2) S_{y_2} + w_3 S_{y_2(2)}, \]

\[ f_i = \left( 1 - \frac{1}{n_i} \right), \quad w_i = \frac{W_i (k-1)}{n_i^2}, \quad W_i = \frac{N_i}{N}, \quad i = 1, 2, 3. \]
Differentiating the above equation, we get the following values

\[ \alpha_1 = \frac{S^*_{yx_1}}{S^2_{x_1}} = \alpha_0 \text{ (say)} , \quad \alpha_2 = \frac{S^*_{yx_2}}{S^2_{x_2}} = \alpha_{00} \text{ (say)} \]

Then we get the optimum value

\[ MSE\left( \hat{t}_{2(11)} \right)_{opt} = S^2_y \left( 1 - \rho^*_{yx_1} - \rho^*_{yx_2} \right) \quad \text{(3.13)} \]

where, \( \rho^*_{yx_1} = \frac{S^*_{yx_1}}{S^2_y S^2_{x_1}} \), \( \rho^*_{yx_2} = \frac{S^*_{yx_2}}{S^2_y S^2_{x_2}} \)

**Efficiency Comparisons:**

In this section, the conditions for which the proposed estimators are better than the usual unbiased estimator \( \bar{y}^* \) have been obtained.

\[ Var\left( \bar{y}^* \right) = \left[ \left( \frac{1}{n} - \frac{1}{N} \right) S^2_y + \frac{W_2 (k-1)}{n} S^2_{y(2)} \right] = f_3 S^2_y + w_3 S^2_{y(2)} = S^2_y \quad \text{(4.1)} \]

\[ MSE\left( \hat{T}_{11} \right)_{min} = S^2_y \left( 1 - R^2_{y,x_1,x_2,x_3,...,x_k} \right) \quad \text{(2.8)} \]

\[ MSE\left( \hat{t}_{k(11)} \right)_{opt} = S^2_y \left( 1 - \sum_{i=1}^{k} \rho^*_{yx_i} \right) \quad \text{(3.8)} \]

Using (4.1) and (2.8), we obtain

\[ Var\left( \bar{y}^* \right) - MSE\left( \hat{T}_{11} \right)_{min} \geq 0 \]

\[ R^2_{y,x_1,x_2,x_3,...,x_k} \geq 0 \]

Using (4.1) and (3.8), we obtain

\[ \sum_{i=1}^{k} \rho^*_{yx_i} \geq 0 \]

**Numerical Illustration:**

To get a rough idea about gain in efficiency of the proposed estimators over the usual unbiased estimator in multi phase sampling scheme, we take the population from Gujarati, D.N. (2003), the variables are defined and values of population parameters are shown. The Percent Relative Efficiency (PRE) for both proposed estimators w.r.t usual unbiased estimator is also shown in Table 1 and Table 2 for different values of \( k \). We also showed results using different values of \( (n_1, n_2, n_3) \).
Let, \( y \) be the average miles per gallons, 
\( x_1 \) be the engine horse powers and 
\( x_2 \) be the cubic feet of cab space.

\[
\begin{align*}
N & = 81, \quad N_1 = 40, \quad N_2 = 35, \quad N_3 = 25, \\
\bar{y} & = 33.82, \quad \bar{x}_1 = 117.47, \quad \bar{x}_2 = 108.79, \quad \rho_{x_1x_2} = 0.62, \\
S^2_y & = 100.55, \quad S^2_{x_1} = 3261.95, \quad S^2_{x_2} = 571.49, \quad S_{x_1x_2} = 833.69, \\
S_{y_{x_1}} & = -449.80, \quad S_{y_{x_2}} = -106.20, \quad \rho_{y_{x_1}} = -0.80, \quad \rho_{y_{x_2}} = -0.45
\end{align*}
\]

We used the following expressions for comparison

\[
\begin{align*}
\text{PRE}(\hat{T}_{11}, \bar{y}^*) & = \frac{\text{Var}(\bar{y}^*)}{\text{Var}(\hat{T}_{11})} \times 100 \\
\text{PRE}(\hat{i}_{k(11)}, \bar{y}^*) & = \frac{\text{Var}(\bar{y}^*)}{\text{Var}(\hat{i}_{k(11)})} \times 100, \quad \text{Var}(\bar{y}^*) = f_s S^2_y + w_s S^2_{y(2)}
\end{align*}
\]

<table>
<thead>
<tr>
<th>( n_1 )</th>
<th>( n_2 )</th>
<th>( n_3 )</th>
<th>((1/k))</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>((1/5)) ((1/4)) ((1/3)) ((1/2))</td>
</tr>
<tr>
<td>49</td>
<td>37</td>
<td>21</td>
<td>137.1214</td>
</tr>
<tr>
<td></td>
<td>17</td>
<td>112.2789</td>
<td>110.8621</td>
</tr>
<tr>
<td>33</td>
<td>13</td>
<td>118.6064</td>
<td>116.7098</td>
</tr>
<tr>
<td></td>
<td>9</td>
<td>111.5639</td>
<td>110.4225</td>
</tr>
<tr>
<td>45</td>
<td>33</td>
<td>21</td>
<td>122.0221</td>
</tr>
<tr>
<td></td>
<td>17</td>
<td>121.2588</td>
<td>118.8647</td>
</tr>
<tr>
<td></td>
<td>13</td>
<td>121.0361</td>
<td>118.8126</td>
</tr>
<tr>
<td></td>
<td>9</td>
<td>112.9733</td>
<td>111.6551</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>( n_1 )</th>
<th>( n_2 )</th>
<th>( n_3 )</th>
<th>((1/k))</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>((1/5)) ((1/4)) ((1/3)) ((1/2))</td>
</tr>
<tr>
<td>49</td>
<td>37</td>
<td>21</td>
<td>168.3358</td>
</tr>
<tr>
<td></td>
<td>17</td>
<td>148.8767</td>
<td>141.2827</td>
</tr>
<tr>
<td>33</td>
<td>13</td>
<td>136.7145</td>
<td>130.7709</td>
</tr>
<tr>
<td></td>
<td>9</td>
<td>125.2077</td>
<td>121.0404</td>
</tr>
<tr>
<td>45</td>
<td>33</td>
<td>21</td>
<td>173.7745</td>
</tr>
<tr>
<td></td>
<td>17</td>
<td>152.1141</td>
<td>143.5454</td>
</tr>
<tr>
<td></td>
<td>13</td>
<td>139.251</td>
<td>132.5727</td>
</tr>
<tr>
<td></td>
<td>9</td>
<td>126.6032</td>
<td>122.0526</td>
</tr>
</tbody>
</table>
CONCLUSION

From Table 1 and Table 2, it is observed that estimators $\hat{T}_{11}$ and $\hat{t}_{k(1)}$ both are more efficient than the usual estimator $\bar{y}^*$ under fixed sample size. As $n_i$ decreases the PRE of proposed estimators increases. Also as the value of $k$ increases, the PRE’s of proposed estimators decrease. To estimate the population mean $\bar{Y}$ of the study variable $y$ in the presence of non-response(on the auxiliary variable also), we take two situations viz. (a) Population mean $\bar{X}$ is known, and (b) population mean $\bar{X}$ is unknown. We proof that proposed estimators are more efficient and should work very well practically.

REFERENCES

A GENERAL FAMILY OF ESTIMATORS WHEN THE STUDY VARIABLE IS AN ATTRIBUTE

Ayesha Tahira\textsuperscript{1} and Javid Shabbir\textsuperscript{2}
Department of Statistics, Quaid-i-Azam University, Islamabad, Pakistan.
Emails: aisha_tahira1@yahoo.com and js_qau@yahoo.com

ABSTRACT

We have developed a family of exponential ratio type estimators when study variable is qualitative in nature, utilizing information on the auxiliary variable under simple random and stratified random sampling. The expressions for bias and mean-squared error (MSE) up to the first order of approximation are derived. Estimators based on estimated optimum values are also considered with their properties. We compare the proposed estimators with the usual proportion estimator, traditional ratio estimator and exponential ratio type estimator and Singh et al. (2010) estimators. We accomplish the efficiency conditions and an empirical study is carried out to judge the merits of the suggested estimators. It is shown that the suggested class of estimators is more efficient than the existing estimators.

KEYWORD

Auxiliary Variable; Study attribute; MSE; Stratified Random Sampling; Point bi-
serial Population Correlation Coefficient.

INTRODUCTION

In survey sampling, estimation of population mean has received considerable attention of the researchers. The use of auxiliary information increases the precision of the estimators by taking the advantage of correlation between the study and the auxiliary variable. Neyman (1938) used such information for the first time in estimating the population mean and got some improved results. The stratified random sampling is most effective when variability within the strata is minimized, between the strata is maximized and a variable used for making strata is highly correlated with the study variable. Various ratio and regression type estimators can be improved in the stratified random sampling using the auxiliary variable and some parameters related to auxiliary variable such as coefficient of variation, coefficient of kurtosis, coefficient of skewness, coefficient of correlation. Dalalbehara and Sahoo (1997, 1999) proposed a class of estimators and developed a regression type estimator in stratified random sampling using two auxiliary variables. Kadilar and Cingi (2003, 2005) and Koyuncu and Kadilar (2008, 2009) have utilized the auxiliary information for estimating the population mean in stratified random sampling.

However in many cases, the study variable is available in qualitative in nature. Singh et al. (2010) have proposed ratio estimator, and a family of estimators for estimating the population proportion. In this paper, we propose a general class of estimators for estimating the population proportion in stratified random sampling.
Consider a finite population $U = (u_1, u_2, u_3, ..., u_N)$ of size $N$ and is divided into $L$ strata such that $\sum_{h=1}^{L} N_h = N$, where $N_h$ is size of the $h^{th}$ stratum. We take a random sample of size $n_h$ from a population of size $N_h$ by simple random sampling without replacement (SRSWOR) such that $\sum_{h=1}^{L} n_h = n$, where $n_h$ is the $h^{th}$ stratum sample size. Let $\phi_{hi}$ and $x_{hi}$ be the characteristics of the study attribute ($\phi$) and the auxiliary variable ($x$) on the $i^{th}$ unit of the $h^{th}$ stratum, respectively. Suppose there is complete dichotomy in the population with respect to the presence or absence of an attribute, say $\phi_h$, and it is assumed that $\phi_{hi}$ can take only two possible values 1 and 0 as:

$$\phi_i = \begin{cases} 1, & \text{if } i^{th} \text{ unit of the } h^{th} \text{ stratum possesses an attribute } \phi \\ 0, & \text{otherwise.} \end{cases}$$

Let $P = \sum_{h=1}^{L} W_h P_h$ and $p_{st} = \sum_{h=1}^{L} W_h p_h$, be the population and sample proportions of the units of the study attribute; $P_h = \frac{A_h}{N_h}$ and $p_h = \frac{a_h}{n_h}$, be the population and sample proportion of units in the stratum $h$, where $A_h = \sum_{i=1}^{N_h} \phi_{hi}$ and $a_h = \sum_{i=1}^{n_h} \phi_{hi}$. To estimate the population proportion of the study attribute i.e. $(P)$, we assume that $\bar{X}$ is known.

Let $\rho_{\phi x h} = \frac{S_{\phi x h}}{S_{\phi h} S_{x h}}$ be the point bi-serial population correlation coefficient between the study character ($\phi$) and the auxiliary variable ($x$). Let $S_{x h}^2 = \frac{1}{N_h - 1} \sum_{i=1}^{N_h} (x_{hi} - \bar{X}_h)^2$ and $S_{\phi h}^2 = \frac{1}{N_h - 1} \sum_{i=1}^{N_h} (\phi_{hi} - P_h)^2$ be the population variances of the auxiliary variable and the study character respectively. Let $S_{\phi x h} = \frac{1}{N_h - 1} \sum_{i=1}^{N_h} (\phi_{hi} - P_h)(x_{hi} - \bar{X}_h)$ be the population point bi-covariance between the study character and the auxiliary variable. Let $C_{\phi h} = \frac{S_{\phi h}}{P}$ and $C_{x h} = \frac{S_{x h}}{\bar{X}}$ be the coefficient of variation of $(P)$ and $(x)$ respectively.

The usual unbiased estimator of the population proportion, is given by

$$t_{st} = p_{st}$$

(1)

The variance for the usual estimator is given by
\[ V(t_{st}) = \sum_{h=1}^{L} W_h^2 \lambda_h C_{qh}^2. \]  

(2)

where \( \lambda_h = \left(1 - f_h\right) \) and \( f_h = \frac{n_h}{N_h} \), also \( W_h = \frac{N_h}{N} \) is the known stratum weight.

Singh et al. (2010) estimators in the stratified random sampling is given by

\[ t_{1(st)} = p_{st} \left( \frac{\bar{X}}{\bar{x}_{st}} \right) \]  

(3)

\[ t_{2(st)} = H(p_{st}, u_{st}) , \]  

(4)

where \( u_{st} = \frac{\bar{x}_{st}}{\bar{X}} \) and \( H(p_{st}, u_{st}) \) is a parametric function of \( p_{st} \) and \( u_{st} \) such that \( H(p_{st},1) = P, \forall P \).

The bias and \( MSE \) expressions of \( t_{j(st)} \) \((j = 1,2)\) estimators, to the first order of approximation, are respectively given by

\[ B(t_{1(st)}) \approx P \sum_{h=1}^{L} W_h \lambda_h \left( C_{qh}^2 - \rho_{qhx} C_{qh} C_{xh} \right). \]

\[ MSE(t_{1(st)}) \approx P^2 \sum_{h=1}^{L} W_h^2 \lambda_h \left( C_{qh}^2 + C_{xh}^2 - 2\rho_{qhx} C_{qh} C_{xh} \right). \]  

(5)

\[ B(t_{2(st)}) \approx \sum_{h=1}^{L} W_h \lambda_h \left( P \rho_{qhx} C_{qh} C_{xh} H_3 + C_{xh}^2 H_2 + P^2 C_{qh}^2 H_4 \right). \]

where \( H_2 = \frac{1}{2} \frac{\partial^2 H}{\partial u^2} \bigg|_{p=p,u=1} \), \( H_3 = \frac{1}{2} \frac{\partial^2 H}{\partial p \partial u} \bigg|_{p=p,u=1} \) and \( H_4 = \frac{1}{2} \frac{\partial^2 H}{\partial p^2} \bigg|_{p=p,u=1} \)  

\[ MSE(t_{2(st)}) \approx P^2 \sum_{h=1}^{L} W_h^2 \lambda_h C_{qh}^2 \left(1 - \rho_{qhx}^2 \right). \]  

(6)

**Proposed Estimator**

In this paper, we propose a generalized exponential estimator for improving the efficiency by using auxiliary variable and taking into consideration the point bi-serial correlation between study attribute and the auxiliary variable, which is given by

\[ \hat{P}_{A(st)} = p_{st} \left[ k_1 \left( \frac{\bar{x}}{\bar{x}_{st}} \right)^{\gamma_1} \exp \left( \frac{(a\bar{x} + b) - (a\bar{x}_{st} + b)}{(a\bar{x} + b) + (a\bar{x}_{st} + b)} \right) + k_2 \left( \frac{\bar{x}}{\bar{x}_{st}} \right)^{\gamma_2} \exp \left( \frac{(a\bar{x}_{st} + b) - (a\bar{x} + b)}{(a\bar{x}_{st} + b) + (a\bar{x} + b)} \right) \right], \]

(7)

where \( k_1 \) and \( k_2 \) are unknown constants to be determined under certain criteria and \( \gamma_1 \) and \( \gamma_2 \) are suitably chosen constants.

Let we define
A general family of estimators when the study variable is an attribute.

\[ e_{pst} = \frac{p_{st} - P}{P} \quad \text{and} \quad e_{xst} = \frac{\bar{x}_{st} - \bar{X}}{\bar{X}}, \]

such that \( E(e_{pst}) = E(e_{xst}) = 0, \) \( E(e_{pst}^2) = \sum_{h=1}^{L} W_h^2 \lambda_h C_{qh}^2, \)
\( E(e_{xst}^2) = \sum_{h=1}^{L} W_h^2 \lambda_h C_{xh}^2, \) \( E(e_{xst} e_{pst}) = \sum_{h=1}^{L} W_h^2 \lambda_h P_{qhxh} C_{ph} C_{xh}. \)

**Bias and MSE of \( \hat{P}_{A(st)} \)**

Now rewriting (7) in terms of \( e \)'s, we get
\[
\hat{P}_{A(st)} = P \left( 1 + e_{pst} \right) \left[ k_1 \left( 1 + e_{xst} \right)^{-\gamma_1} \exp \left\{ \frac{-a\bar{X}e_{xst}}{a\bar{X}e_{xst} + 2a\bar{X} + 2b} \right\} \right] + k_2 \left( 1 + e_{xst} \right)^{-\gamma_2} \exp \left\{ \frac{a\bar{X}e_{xst}}{a\bar{X}e_{xst} + 2a\bar{X} + 2b} \right\}
\]

or
\[
\hat{P}_{A(st)} = P \left( 1 + e_{pst} \right) \left[ k_1 \left( 1 + e_{xst} \right)^{-\gamma_1} \exp \left\{ -\frac{\mu e_{xst}}{2} \left( 1 + \frac{\mu e_{xst}}{2} \right)^{-1} \right\} \right] + k_2 \left( 1 + e_{xst} \right)^{-\gamma_2} \exp \left\{ \frac{\mu e_{xst}}{2} \left( 1 + \frac{\mu e_{xst}}{2} \right)^{-1} \right\}
\]
where \( \mu = \frac{a\bar{X}}{a\bar{X} + b} \).

Expanding up to first degree of approximation, we have
\[
\hat{P}_{A(st)} - P = P \left[ k_1 \left\{ 1 + e_{pst} - \frac{\mu + 2\gamma_1}{2} e_{xst} - \frac{\mu + 2\gamma_1}{2} e_{xst} e_{pst} + \frac{\gamma_1 (\gamma_1 + 1)}{2} e_{xst}^2 + \frac{\mu \gamma_1}{2} e_{xst}^2 + \frac{3\mu^2}{8} e_{xst}^2 \right\} \right.
\]
\[
+ k_2 \left\{ 1 + e_{xst} + \frac{\mu - 2\gamma_2}{2} e_{xst} - \frac{\mu - 2\gamma_2}{2} e_{xst} e_{xst} + \frac{\gamma_2 (\gamma_2 + 1)}{2} e_{xst}^2 - \frac{\mu \gamma_2}{2} e_{xst}^2 - \frac{\mu^2}{8} e_{xst}^2 \right\} - 1 \right]
\]
\( (8) \)

**Case I:** when \( k_1 + k_2 = 1 \)
\[
\hat{P}_{A(st)} - P = P \left[ k_1 \left\{ e_{pst} - \frac{\mu + 2\gamma_1}{2} e_{xst} - \frac{\mu + 2\gamma_1}{2} e_{xst} e_{pst} + \frac{\gamma_1 (\gamma_1 + 1)}{2} e_{xst}^2 + \frac{\mu \gamma_1}{2} e_{xst}^2 + \frac{3\mu^2}{8} e_{xst}^2 \right\} \right.
\]
\[
+ (1 - k_1) \left\{ e_{pst} + \frac{\mu - 2\gamma_2}{2} e_{xst} + \frac{\mu - 2\gamma_2}{2} e_{xst} e_{xst} + \frac{\gamma_2 (\gamma_2 + 1)}{2} e_{xst}^2 - \frac{\mu \gamma_2}{2} e_{xst}^2 - \frac{\mu^2}{8} e_{xst}^2 \right\} - 1 \right]
\]
\( (9) \)
Taking expectation of (9), we get bias of \( \hat{P}_{A(st)} \) as given by

\[
Bias\left( \hat{P}_{A(st)} \right) = E\left( \hat{P}_{A(st)} - P \right) = P \sum_{h=1}^{L} \lambda_{h} \left[ \frac{(\mu - 2\gamma_2)}{2} C_{qhx} + \frac{\gamma_2(\gamma_2 + 1)}{2} C_{xh} - \frac{\mu^2}{2} C_{xh}^2 - \frac{\mu^2}{8} C_{xh}^2 \right. \\
+ k_1 \left[ \frac{\gamma_1(\gamma_1 + 1)}{2} C_{xh}^2 - \frac{(\mu + 2\gamma_1)}{2} C_{qhx} - \frac{(\mu - 2\gamma_2)}{2} C_{qhx} \right. \\
\left. - \frac{\gamma_2(\gamma_2 + 1)}{2} C_{xh}^2 + \frac{\mu(\gamma_2 - \gamma_1)}{2} C_{xh}^2 + \frac{\mu^2}{2} C_{xh}^2 \right]. \tag{10}
\]

Using (9), we have

\[
\left( \hat{P}_{A(st)} - P \right)^2 = P^2 \left[ k_1 e_{pst} - k_1 \frac{(\mu + 2\gamma_1)}{2} e_{xst} + e_{pst} + \frac{(\mu - 2\gamma_2)}{2} e_{xst} - k_1 \frac{(\mu - 2\gamma_2)}{2} e_{xst} \right]^2. \tag{11}
\]

or

\[
MSE\left( \hat{P}_{A(st)} \right) = E\left( \hat{P}_{A(st)} - P \right)^2 = P^2 \sum_{h=1}^{L} \lambda_{h} W_h^2 \left[ C_{qhx}^2 + \frac{(\alpha_{st} - 2\gamma_2)^2}{4} C_{xh}^2 \right. \\
+ k_1^2 (\mu + \gamma_1 - \gamma_2)^2 C_{xh}^2 + (\mu - 2\gamma_2) C_{qhx} - 2k_1 (\mu + \gamma_1 - \gamma_2) C_{qhx} \right. \\
\left. - k_1 (\mu - 2\gamma_2)(\mu + \gamma_1 - \gamma_2) C_{xh}^2 \right]. \tag{12}
\]

Minimizing for the value of ‘\( k_1 \)’

\[
\frac{\partial MSE}{\partial k_1} = 2k_1 (\mu + \gamma_1 - \gamma_2)^2 C_{xh}^2 - 2(\mu + \gamma_1 - \gamma_2) C_{qhx} - (\mu - 2\gamma_2)(\mu + \gamma_1 - \gamma_2) C_{xh}^2 = 0
\]

we get

\[
k_1 = \frac{2 \sum_{h=1}^{L} \lambda_{h} W_h^2 C_{qhx} + (\mu - 2\gamma_2) \sum_{h=1}^{L} \lambda_{h} W_h^2 C_{xh}^2}{2(\mu + \gamma_1 - \gamma_2) \sum_{h=1}^{L} \lambda_{h} W_h^2 C_{xh}^2} = k_{opt} \text{ (say)}.
\]

Substituting the optimum value of \( k_1 \) in (3.14), we get the optimal \( MSE(\hat{P}_{A(st)}) \)

which is given by

\[
MSE(\hat{P}_{A(st)})_{opt} = P^2 \sum_{h=1}^{L} W_h \lambda_h C_{qhx}^2 \left( 1 - \rho_{qhx}^2 \right). \tag{13}
\]

which is equal to variance of the linear regression estimator.

**Case II:** when \( k_1 + k_2 \neq 1 \)
\[
\hat{P}_{(st)}^\prime - P = \frac{k_1\left\{ 1 + e_{pst} \left( \frac{\mu + 2\gamma_1}{2} \right) e_{xst} - \frac{\mu + 2\gamma_1}{2} e_{xst} e_{pst} + \frac{\gamma_1 (\gamma_1 + 1)}{2} e_{xst}^2 + \frac{\mu^2 + 3\mu^2}{8} e_{xst}^2 \right\}} + k_2\left\{ 1 + e_{pst} \left( \frac{\mu - 2\gamma_2}{2} \right) e_{xst} - \frac{\mu - 2\gamma_2}{2} e_{xst} e_{pst} + \frac{\gamma_2 (\gamma_2 + 1)}{2} e_{xst}^2 - \frac{\mu^2 + 3\mu^2}{8} e_{xst}^2 \right\} - 1
\]

Using (14), we have

\[
\text{Bias}\left( \hat{P}_{(st)}^\prime \right) = E \left( \hat{P}_{(st)}^\prime - P \right)
\]

\[
\text{Bias}\left( \hat{P}_{(st)}^\prime \right) \equiv P \sum_{h=1}^\ell W_h^2 \lambda_h \left[ k_1 \left\{ \frac{\gamma_1 (\gamma_1 + 1)}{2} C_{xh}^2 - \frac{\mu + 2\gamma_1}{2} C_{xh} + \frac{\mu^2 + 3\mu^2}{8} C_{xh} \right\} + k_2 \left\{ \frac{\mu - 2\gamma_2}{2} C_{xh} + \frac{\gamma_2 (\gamma_2 + 1)}{2} C_{xh} - \frac{\mu^2 + 3\mu^2}{8} C_{xh} \right\} - 1 \right]
\]

Using (14), we have

\[
\text{MSE}\left( \hat{P}_{(st)}^\prime \right) \equiv P^2 \left[ 1 + A_{st} k_1^2 + B_{st} k_2^2 + 2k_1 k_2 C_{st} - 2k_1 D_{st} - 2k_2 E_{st} \right]
\]

where

\[
A_{st} = 1 + \sum_{h=1}^\ell W_h^2 \lambda_h \left[ C_{qh}^2 - 2(\mu + 2\gamma_1) C_{xh} + \frac{(\mu + 2\gamma_1)^2}{4} C_{xh} + \gamma_1 (\gamma_1 + 1) C_{xh}^2 + \mu \gamma_1 C_{xh}^2 + \frac{3\mu^2}{4} C_{xh} \right]
\]

\[
B_{st} = 1 + \sum_{h=1}^\ell W_h^2 \lambda_h \left[ C_{qh}^2 + \frac{(\mu - 2\gamma_2)^2}{2} C_{xh}^2 + 4(\mu - 2\gamma_2) C_{xh} + \gamma_2 (\gamma_2 + 1) C_{xh}^2 - \mu \gamma_2 C_{xh}^2 - \frac{\mu^2}{4} C_{xh} \right]
\]

\[
C_{st} = 1 + \sum_{h=1}^\ell W_h^2 \lambda_h \left[ C_{qh}^2 - \frac{\gamma_1 (\gamma_1 + 1)}{2} C_{xh}^2 + \frac{\gamma_2 (\gamma_2 + 1)}{2} C_{xh}^2 - (\mu + 2\gamma_1) C_{xh} + \frac{3(\mu - 2\gamma_2)}{2} C_{xh} \right]
\]

\[
D_{st} = 1 - \sum_{h=1}^\ell W_h^2 \lambda_h \left[ \frac{(\mu + 2\gamma_1)}{2} C_{xh} + \frac{\gamma_1 (\gamma_1 + 1)}{2} C_{xh}^2 + \frac{\mu \gamma_1}{2} C_{xh}^2 + \frac{3\mu^2}{8} C_{xh} \right]
\]

and

\[
E_{st} = 1 + \sum_{h=1}^\ell W_h^2 \lambda_h \left[ \frac{(\mu - 2\gamma_2)}{2} C_{xh} + \frac{\gamma_2 (\gamma_2 + 1)}{2} C_{xh}^2 - \frac{\mu \gamma_2}{2} C_{xh}^2 - \frac{\mu^2}{8} C_{xh} \right]
\]

Now minimizing the (15) for the values of \( k_1 \) and \( k_2 \), we get

\[
k_{1(\text{opt})} = \frac{B_{st} D_{st} - E_{st} C_{st}}{A_{st} E_{st} - C_{st} D_{st}}, \quad k_{2(\text{opt})} = \frac{A_{st} E_{st} - C_{st} D_{st}}{A_{st} B_{st} - C_{st}^2}.
\]
Substituting the optimum value of \( k_{1(\text{opt})} \) and \( k_{2(\text{opt})} \) in (15), we get the optimum

\[
\text{MSE}(\hat{P}_{A(st)}) \approx p^2 \left[ 1 - \frac{B_{st}D_{st}^2 + A_{st}E_{st}^2}{A_{st}B_{st} - C_{st}^2} \right]
\]

(16)

Some Members of the Suggested Family of Estimators

Some members of the suggested family of estimators for \( \gamma_1 = \gamma_2 = 1 \) by substituting different choices of \( a \) and \( b \) in (7).

Putting \( a = 1, b = C_x \) in (7), we have

\[
\hat{P}_{R_{st}} = p \left[ k_1 \left( \frac{x}{x_{st}} \right) \exp \left\{ \frac{\bar{X} + C_x}{\bar{X} + C_x + \bar{x}_{st} + C_x} + k_2 \left( \frac{x}{x_{st}} \right) \exp \left\{ \frac{\bar{X} + C_x}{\bar{X} + C_x + \bar{x}_{st} + C_x} \right\} \right] \right]
\]

Putting \( a = 1, b = \rho_{qx} \)

\[
\hat{P}_{R_{st}} = p \left[ k_1 \left( \frac{x}{x_{st}} \right) \exp \left\{ \frac{\bar{X} + \rho_{qx}}{\bar{X} + \rho_{qx} + \bar{x}_{st} + \rho_{qx}} + k_2 \left( \frac{x}{x_{st}} \right) \exp \left\{ \frac{\bar{X} + \rho_{qx}}{\bar{X} + \rho_{qx} + \bar{x}_{st} + \rho_{qx}} \right\} \right] \right]
\]

Putting \( a = C_x, b = \rho_{qx} \)

\[
\hat{P}_{R_{st}} = p \left[ k_1 \left( \frac{x}{x_{st}} \right) \exp \left\{ \frac{\bar{X} + \rho_{qx} \bar{x}_{st} + \rho_{qx}}{\bar{X} + \rho_{qx} \bar{x}_{st} + \rho_{qx} + \bar{x}_{st} + \rho_{qx}} + k_2 \left( \frac{x}{x_{st}} \right) \exp \left\{ \frac{\bar{X} + \rho_{qx} \bar{x}_{st} + \rho_{qx}}{\bar{X} + \rho_{qx} \bar{x}_{st} + \rho_{qx} + \bar{x}_{st} + \rho_{qx}} \right\} \right] \right]
\]

Putting \( a = \rho_{qx}, b = C_x \)

\[
\hat{P}_{R_{st}} = p \left[ k_1 \left( \frac{x}{x_{st}} \right) \exp \left\{ \frac{\rho_{qx} \bar{X} + C_x}{\rho_{qx} \bar{X} + C_x + \bar{x}_{st} + C_x} + k_2 \left( \frac{x}{x_{st}} \right) \exp \left\{ \frac{\rho_{qx} \bar{X} + C_x}{\rho_{qx} \bar{X} + C_x + \bar{x}_{st} + C_x} \right\} \right] \right]
\]

Some members of the suggested family of estimators for \( \gamma_1 = \gamma_2 = -1 \).

Putting \( a = 1, b = C_x \)

\[
\hat{P}_{R_{st}} = p \left[ k_1 \left( \frac{x_{st}}{x} \right) \exp \left\{ \frac{\bar{X} + C_x}{\bar{X} + C_x + \bar{x}_{st} + C_x} + k_2 \left( \frac{x_{st}}{x} \right) \exp \left\{ \frac{\bar{X} + C_x}{\bar{X} + C_x + \bar{x}_{st} + C_x} \right\} \right] \right]
\]
Putting \( a = 1, \ b = \rho_{qx} \)

\[
\hat{P}_{xst} = p \left[ k_1 \left( \frac{\bar{x}_{st}}{X} \right) \exp \left\{ \left( \frac{X + \rho_{qx}}{X} \right) - \left( \frac{\bar{x}_{st} + \rho_{qx}}{\bar{x}_{st} + \rho_{qx}} \right) \right\} + k_2 \left( \frac{\bar{x}_{st}}{X} \right) \exp \left\{ \left( \frac{\bar{x}_{st} + \rho_{qx}}{\bar{x}_{st} + \rho_{qx}} \right) - \left( \frac{X + \rho_{qx}}{X} \right) \right\} \right]
\]

Putting \( a = C_x, \ b = \rho_{qx} \)

\[
\hat{P}_{xst} = p \left[ k_1 \left( \frac{\bar{x}_{st}}{X} \right) \exp \left\{ \left( \frac{C_x \bar{X} + \rho_{qx}}{C_x \bar{X} + \rho_{qx}} \right) - \left( \frac{C_x \bar{x}_{st} + \rho_{qx}}{C_x \bar{x}_{st} + \rho_{qx}} \right) \right\} + k_2 \left( \frac{\bar{x}_{st}}{X} \right) \exp \left\{ \left( \frac{C_x \bar{x}_{st} + \rho_{qx}}{C_x \bar{x}_{st} + \rho_{qx}} \right) - \left( \frac{C_x \bar{X} + \rho_{qx}}{C_x \bar{X} + \rho_{qx}} \right) \right\} \right]
\]

Putting \( a = \rho_{qx}, \ b = C_x \)

\[
\hat{P}_{xst} = p \left[ k_1 \left( \frac{\bar{x}_{st}}{X} \right) \exp \left\{ \left( \frac{\rho_{qx} \bar{X} + C_x}{\rho_{qx} \bar{X} + C_x} \right) - \left( \frac{\rho_{qx} \bar{x}_{st} + C_x}{\rho_{qx} \bar{x}_{st} + C_x} \right) \right\} + k_2 \left( \frac{\bar{x}_{st}}{X} \right) \exp \left\{ \left( \frac{\rho_{qx} \bar{x}_{st} + C_x}{\rho_{qx} \bar{x}_{st} + C_x} \right) - \left( \frac{\rho_{qx} \bar{X} + C_x}{\rho_{qx} \bar{X} + C_x} \right) \right\} \right]
\]

Some members of the suggested family of estimators for \( \gamma_1 = \gamma_2 = 0 \).

Putting \( a = 1, \ b = C_x \)

\[
\hat{P}_{\text{exp,} \ xst} = p \left[ k_1 \exp \left\{ \left( \frac{\bar{X} + C_x}{\bar{X} + C_x} \right) - \left( \frac{\bar{x}_{st} + C_x}{\bar{x}_{st} + C_x} \right) \right\} + k_2 \exp \left\{ \left( \frac{\bar{x}_{st} + C_x}{\bar{x}_{st} + C_x} \right) - \left( \frac{\bar{X} + C_x}{\bar{X} + C_x} \right) \right\} \right]
\]

Putting \( a = 1, \ b = \rho_{qx} \)

\[
\hat{P}_{\text{exp,} \ xst} = p \left[ k_1 \exp \left\{ \left( \frac{\bar{X} + \rho_{qx}}{\bar{X} + \rho_{qx}} \right) - \left( \frac{\bar{x}_{st} + \rho_{qx}}{\bar{x}_{st} + \rho_{qx}} \right) \right\} + k_2 \exp \left\{ \left( \frac{\bar{x}_{st} + \rho_{qx}}{\bar{x}_{st} + \rho_{qx}} \right) - \left( \frac{\bar{X} + \rho_{qx}}{\bar{X} + \rho_{qx}} \right) \right\} \right]
\]

Putting \( a = C_x, \ b = \rho_{qx} \)

\[
\hat{P}_{\text{exp,} \ xst} = p \left[ k_1 \exp \left\{ \left( \frac{C_x \bar{X} + \rho_{qx}}{C_x \bar{X} + \rho_{qx}} \right) - \left( \frac{C_x \bar{x}_{st} + \rho_{qx}}{C_x \bar{x}_{st} + \rho_{qx}} \right) \right\} + k_2 \exp \left\{ \left( \frac{C_x \bar{x}_{st} + \rho_{qx}}{C_x \bar{x}_{st} + \rho_{qx}} \right) - \left( \frac{C_x \bar{X} + \rho_{qx}}{C_x \bar{X} + \rho_{qx}} \right) \right\} \right]
\]

Putting \( a = \rho_{qx}, \ b = C_x \)

\[
\hat{P}_{\text{exp,} \ xst} = p \left[ k_1 \exp \left\{ \left( \frac{\rho_{qx} \bar{X} + C_x}{\rho_{qx} \bar{X} + C_x} \right) - \left( \frac{\rho_{qx} \bar{x}_{st} + C_x}{\rho_{qx} \bar{x}_{st} + C_x} \right) \right\} + k_2 \exp \left\{ \left( \frac{\rho_{qx} \bar{x}_{st} + C_x}{\rho_{qx} \bar{x}_{st} + C_x} \right) - \left( \frac{\rho_{qx} \bar{X} + C_x}{\rho_{qx} \bar{X} + C_x} \right) \right\} \right]
\]
Bias and MSE of the Given Estimators When \( k_1 + k_2 = 1 \)

Expressions for bias of \( \hat{P}_{A, A'_n} (i = 1, 2, \ldots, 4) \), to the first order of approximation, are given by

For \( \gamma_1 = \gamma_2 = 1 \)

\[
Bias(\hat{P}_{R, A'_n}) \approx P \sum_{h=1}^{L} \lambda_h W_h^2 \left[ \frac{\mu_i - 2\gamma_2}{2} C_{qhx} + \frac{\gamma_2 (\gamma_2 + 1)}{2} C_{xh} - \frac{\mu_i \gamma_2}{2} C_{xh} - \frac{\mu_i^2}{8} C_{xh} \right] + k_1 \left[ \frac{\gamma_1 (\gamma_1 + 1)}{2} C_{xh} - \frac{\mu_i + 2\gamma_1}{2} C_{qhx} - \frac{\mu_i - 2\gamma_2}{2} C_{qhx} \right. \\
\left. - \frac{\gamma_2 (\gamma_2 + 1)}{2} C_{xh} + \frac{\mu_i (\gamma_2 - \gamma_1)}{2} C_{xh} + \frac{\mu_i^2}{2} C_{xh} \right].
\]

where \( i = 1, 2, 3, \ldots, 4 \).

For \( \gamma_1 = \gamma_2 = -1 \)

\[
Bias(\hat{P}_{P, A'_n}) \approx P \sum_{h=1}^{L} \lambda_h W_h^2 \left[ \frac{\mu_i - 2\gamma_2}{2} C_{qhx} + \frac{\gamma_2 (\gamma_2 + 1)}{2} C_{xh} - \frac{\mu_i \gamma_2}{2} C_{xh} - \frac{\mu_i^2}{8} C_{xh} \right] + k_1 \left[ \frac{\gamma_1 (\gamma_1 + 1)}{2} C_{xh} - \frac{\mu_i + 2\gamma_1}{2} C_{qhx} - \frac{\mu_i - 2\gamma_2}{2} C_{qhx} \right. \\
\left. - \frac{\gamma_2 (\gamma_2 + 1)}{2} C_{xh} + \frac{\mu_i (\gamma_2 - \gamma_1)}{2} C_{xh} + \frac{\mu_i^2}{2} C_{xh} \right].
\]

where \( i = 1, 2, 3, \ldots, 4 \).

For \( \gamma_1 = \gamma_2 = 0 \)

\[
Bias(\hat{P}_{exp, A'_n}) \approx P \sum_{h=1}^{L} \lambda_h W_h^2 \left[ \frac{\mu_i - 2\gamma_2}{2} C_{qhx} + \frac{\gamma_2 (\gamma_2 + 1)}{2} C_{xh} - \frac{\mu_i \gamma_2}{2} C_{xh} - \frac{\mu_i^2}{8} C_{xh} \right] + k_1 \left[ \frac{\gamma_1 (\gamma_1 + 1)}{2} C_{xh} - \frac{\mu_i + 2\gamma_1}{2} C_{qhx} - \frac{\mu_i - 2\gamma_2}{2} C_{qhx} \right. \\
\left. - \frac{\gamma_2 (\gamma_2 + 1)}{2} C_{xh} + \frac{\mu_i (\gamma_2 - \gamma_1)}{2} C_{xh} + \frac{\mu_i^2}{2} C_{xh} \right].
\]

where \( i = 1, 2, 3, \ldots, 4 \).

and \( \mu_1 = \frac{\bar{X}}{\bar{X} + C_x}, \mu_2 = \frac{\bar{X}}{\bar{X} + \rho_{qx}}, \mu_3 = \frac{C_x \bar{X}}{C_x \bar{X} + \rho_{qx}}, \mu_4 = \frac{\rho_{qx} \bar{X}}{\rho_{qx} \bar{X} + C_x}. \)

MSE expressions of \( \hat{P}_{A_i(s)} (i = 1, 2, \ldots, 4) \), to first order approximation are given by
A general family of estimators when the study variable is an attribute

\[ \text{MSE}(\hat{P}_{A(st)})_{opt} \geq P^2 \sum_{h=1}^{L} W_h^2 \lambda_h C_{qhh} \left( 1 - \rho_{qhx(st)}^2 \right), \]  

(17)

where \( \rho_{qhx(st)} = \frac{\sum_{h=1}^{L} W_h^2 \lambda_h C_{qhh}}{\sqrt{\sum_{h=1}^{L} W_h^2 \lambda_h C_{qhh}^2 \sum_{h=1}^{L} W_h^2 \lambda_h C_{xh}^2}} \)

**Bias and MSE of the Given Estimators When** \( k_1 + k_2 \neq 1 \)

Expressions for the bias of the members of the suggested family of estimators, to the first order of approximation, are given by

For \( \gamma_1 = \gamma_2 = 1 \)

\[ \text{Bias}(\hat{P}_{R(st)}) \equiv P \sum_{h=1}^{L} W_h^2 \lambda_h \left[ k_1 \left\{ \frac{\gamma_1 (\gamma_1 + 1)}{2} C_{xh} + \frac{\mu_i + 2 \gamma_1}{2} C_{qhx} + \frac{\mu_i \gamma_1}{2} C_{xh} + \frac{3 \mu_i^2}{8} C_{xh} \right\} \right. \]

\[ + k_2 \left\{ \frac{(\mu_i - 2 \gamma_2)}{2} C_{qhx} + \frac{\gamma_2 (\gamma_2 + 1)}{2} C_{xh} - \frac{\mu_i \gamma_2}{2} C_{xh} - \frac{\mu_i^2}{8} C_{xh} \right\} \]  

where \( i = 1, 2, 3, \ldots, 4 \).

For \( \gamma_1 = \gamma_2 = -1 \)

\[ \text{Bias}(\hat{P}_{r(st)}) \equiv P \sum_{h=1}^{L} W_h^2 \lambda_h \left[ k_1 \left\{ \frac{\gamma_1 (\gamma_1 + 1)}{2} C_{xh} - \frac{\mu_i + 2 \gamma_1}{2} C_{qhx} + \frac{\mu_i \gamma_1}{2} C_{xh} + \frac{3 \mu_i^2}{8} C_{xh} \right\} \right. \]

\[ + k_2 \left\{ \frac{(\mu_i - 2 \gamma_2)}{2} C_{qhx} + \frac{\gamma_2 (\gamma_2 + 1)}{2} C_{xh} - \frac{\mu_i \gamma_2}{2} C_{xh} - \frac{\mu_i^2}{8} C_{xh} \right\} \]  

where \( i = 1, 2, 3, \ldots, 4 \).

For \( \gamma_1 = \gamma_2 = 0 \)

\[ \text{Bias}(\hat{P}_{\exp(st)}) \equiv P \sum_{h=1}^{L} W_h^2 \lambda_h \left[ k_1 \left\{ \frac{\gamma_1 (\gamma_1 + 1)}{2} C_{xh} - \frac{\mu_i + 2 \gamma_1}{2} C_{qhx} + \frac{\mu_i \gamma_1}{2} C_{xh} + \frac{3 \mu_i^2}{8} C_{xh} \right\} \right. \]

\[ + k_2 \left\{ \frac{(\mu_i - 2 \gamma_2)}{2} C_{qhx} + \frac{\gamma_2 (\gamma_2 + 1)}{2} C_{xh} - \frac{\mu_i \gamma_2}{2} C_{xh} - \frac{\mu_i^2}{8} C_{xh} \right\} \]  

where \( i = 1, 2, 3, \ldots, 4 \).

**MSE expressions of** \( \hat{P}_{A(st)} \) \( (i = 1, 2, \ldots, 4) \), to first order approximation are given by

\[ \text{MSE}(\hat{P}_{A(st)})_{opt} \equiv P^2 \left[ 1 - \frac{B_{ist} D_{ist}^2 + A_{ist} E_{ist}^2 - 2 C_{ist} D_{ist} E_{ist}}{A_{ist} B_{ist} - C_{ist}^2} \right] \]  

(18)
where

\[ A_{ist} = 1 + \sum_{h=1}^{L} W_h^2 \lambda_h \left[ C_{qh}^2 - 2(\mu_i + 2\gamma_1)C_{qhx} + \frac{(\mu_i + 2\gamma_1)^2}{4} - C_{xh}^2 + \gamma_1(\gamma_1 + 1)C_{xh}^2 + \mu_i\gamma_1C_{xh}^2 + \frac{3\mu_i^2}{4}C_{xh}^2 \right] \]

\[ B_{ist} = 1 + \sum_{h=1}^{L} W_h^2 \lambda_h \left[ C_{qh}^2 + \frac{(\mu_i - 2\gamma_2)^2}{2}C_{xh}^2 + 4(\mu_i - 2\gamma_2)C_{qhx} + \gamma_2(\gamma_2 + 1)C_{xh}^2 - \mu_i\gamma_2C_{xh}^2 - \frac{\mu_i^2}{4}C_{xh}^2 \right] \]

\[ C_{ist} = 1 + \sum_{h=1}^{L} W_h^2 \lambda_h \left[ C_{qh}^2 - \frac{\gamma_1(\gamma_1 + 1)}{2}C_{xh}^2 + \frac{\gamma_2(\gamma_2 + 1)}{2}C_{xh}^2 - (\mu_i + 2\gamma_1)C_{qhx} + \frac{3(\mu_i - 2\gamma_2)}{2}C_{qhx} \right. \\
\left. \quad - \frac{(\mu_i + 2\gamma_1)(\mu_i - 2\gamma_2)}{4}C_{xh}^2 + \frac{\mu_i(\gamma_1 - \gamma_2)}{2}C_{xh}^2 + \frac{\mu_i^2}{4}C_{xh}^2 \right] \]

\[ D_{ist} = 1 - \sum_{h=1}^{L} W_h^2 \lambda_h \left[ \frac{(\mu_i + 2\gamma_1)}{2}C_{qhx} + \frac{\gamma_1(\gamma_1 + 1)}{2}C_{xh}^2 + \frac{\mu_i\gamma_1}{2}C_{xh}^2 + \frac{3\mu_i^2}{8}C_{xh}^2 \right] \]

\[ E_{ist} = 1 + \sum_{h=1}^{L} W_h^2 \lambda_h \left[ \frac{(\mu_i - 2\gamma_2)}{2}C_{qhx} + \frac{\gamma_2(\gamma_2 + 1)}{2}C_{xh}^2 - \frac{\mu_i\gamma_2}{2}C_{xh}^2 - \frac{\mu_i^2}{8}C_{xh}^2 \right] \]

A Comparative Study of Proposed Estimator with the Various Existing Estimators

To compare the efficiencies of various existing estimators with that of the proposed optimal estimator, we require their mean square errors. So the comparisons of the mean square errors, up to first order of approximation, of various existing estimators considered in this paper are given as under.

Comparison with simple mean per unit estimator

(i) By (2) and (13)

\[ MSE(\hat{P}_{A(st)})_{opt} \leq V(t) , \text{ if} \]

\[ V(t) - MSE(\hat{P}_A)_{opt} \geq 0 , \text{ or if} \]

\[ \rho_{qx(st)}^2 \geq \frac{P^2 - 1}{P^2} . \quad (19) \]

(ii) By (5) and (13)

\[ MSE(\hat{P}_{A(st)})_{opt} \leq MSE(t_{1(st)}) , \text{ if} \]

\[ MSE(t_{1(st)}) - MSE(\hat{P}_A)_{opt} \geq 0 , \text{ or if} \]

\[ \rho_{qx(st)} \geq \frac{\sum_{h=1}^{L} W_h^2 \lambda_h C_{xh}}{\sum_{h=1}^{L} W_h^2 \lambda_h C_{qh}} . \quad (20) \]
(iii) By (6) and (13)

\[ \text{MSE}(\hat{P}_{A(st)})_{opt} = \text{MSE}(t_{2(st)}) . \]

(iv) By (2) and (16)

\[ \text{MSE}(\hat{P}_{A(st)})_{opt} \leq V(t_{st}) , \text{ if } \]

\[ V(t_{st}) - \text{MSE}(\hat{P}_{A(st)})_{opt} \geq 0 , \text{ or if } \]

\[ P^2 \left[ 1 - \frac{B_{st}D_{st}^2 + A_{st}E_{st}^2 - 2C_{st}D_{st}E_{st}}{A_{st}B_{st} - C_{st}^2} \right] \leq \rho_{q(x(st))}^2 . \quad (21) \]

(v) By (5) and (16)

\[ \text{MSE}(\hat{P}_{A(st)})_{opt} \leq \text{MSE}(t_{l(st)}) , \text{ if } \]

\[ \text{MSE}(t_{l(st)}) - \text{MSE}(\hat{P}_{A(st)})_{opt} \geq 0 , \text{ or if } \]

\[ \left[ 1 - \frac{B_{st}D_{st}^2 + A_{st}E_{st}^2 - 2C_{st}D_{st}E_{st}}{A_{st}B_{st} - C_{st}^2} \right] \leq \sum_{h=1}^{L} W_h^2 \lambda_h \left( C_{qhx}^2 + C_{xh}^2 - 2\rho_{qhx}C_{qhx}C_{xh} \right) . \quad (22) \]

(vi) By (6) and (16)

\[ \text{MSE}(\hat{P}_{A(st)})_{opt} \leq \text{MSE}(t_{2(st)}) , \text{ if } \]

\[ \text{MSE}(t_{2(st)}) - \text{MSE}(\hat{P}_{A(st)})_{opt} \geq 0 , \text{ or if } \]

\[ \left[ 1 - \frac{B_{st}D_{st}^2 + A_{st}E_{st}^2 - 2C_{st}D_{st}E_{st}}{A_{st}B_{st} - C_{st}^2} \right] \leq \sum_{h=1}^{L} W_h^2 \lambda_h C_{qhx}^2 \left( 1 - \rho_{q(x(st))}^2 \right) . \quad (23) \]

**Numerical analysis**

We consider the following population to illustrate the properties of the proposed estimator of the population proportion \( P \). For this population we select a random samples of sizes \( n_h \) from each stratum and the sample sizes are computed using Neyman allocation method.

Population: source [Murthy (1967)]

The population consists of village wise complete enumeration data obtained in 1951 and 1961 censuses for a tehsil. The variable area of village is used to stratify the population into 3 strata.

Let \( x \) be the cultivated area in the village in hectares in 1951.

Let \( \varphi_{1i} = \begin{cases} 1, & \text{if } i^{th} \text{ village in the Stratum 1 has an area greater than 550 hectares.} \\ 0, & \text{otherwise.} \end{cases} \)

\[ \varphi_{1i} = \begin{cases} 1, & \text{if } i^{th} \text{ village in the Stratum 2 has an area greater than 1300 hectares.} \\ 0, & \text{otherwise.} \end{cases} \]
and,  \( \phi_{3i} = \begin{cases} 1, & \text{if } i^{th} \text{ village in the Stratum 3 has an area greater than 2500 hectares.} \\ 0, & \text{otherwise.} \end{cases} \)

Data set is presented in Table 1 and results are given in Table 2.

### Table 1: Data Statistic for Population 1

<table>
<thead>
<tr>
<th>Stratum number</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>( N_h )</td>
<td>43</td>
<td>45</td>
<td>40</td>
</tr>
<tr>
<td>( n_h )</td>
<td>10</td>
<td>12</td>
<td>18</td>
</tr>
<tr>
<td>( \bar{X}_h )</td>
<td>397.1425</td>
<td>760.1795</td>
<td>1234.6180</td>
</tr>
<tr>
<td>( P_h )</td>
<td>0.5814</td>
<td>0.4444</td>
<td>0.4000</td>
</tr>
<tr>
<td>( S_{xh}^2 )</td>
<td>39975.0569</td>
<td>61455.9900</td>
<td>172425.9000</td>
</tr>
<tr>
<td>( S_{\phi h}^2 )</td>
<td>0.2492</td>
<td>0.2525</td>
<td>0.2462</td>
</tr>
<tr>
<td>( \rho_{\phi xh} )</td>
<td>0.6922</td>
<td>0.3750</td>
<td>0.5057</td>
</tr>
</tbody>
</table>

### Table 2:

Percent relative efficiency (PRE) of different estimators using auxiliary variable

<table>
<thead>
<tr>
<th></th>
<th>MSE</th>
<th>R.E</th>
</tr>
</thead>
<tbody>
<tr>
<td>( t )</td>
<td>0.020633</td>
<td>100</td>
</tr>
<tr>
<td>( t_1 )</td>
<td>0.003506</td>
<td>588.4947</td>
</tr>
<tr>
<td>( t_2 )</td>
<td>0.003447</td>
<td>598.5355</td>
</tr>
<tr>
<td>( \hat{P}_{A(st)} )</td>
<td>0.003447</td>
<td>598.5355</td>
</tr>
<tr>
<td>( \hat{P}_{R(st)}^{*} )</td>
<td>0.000802</td>
<td>2571.542</td>
</tr>
<tr>
<td>( \hat{P}<em>{R</em>{1}(st)}^{*} )</td>
<td>0.001454</td>
<td>1418.866</td>
</tr>
<tr>
<td>( \hat{P}<em>{A</em>{1}(st)} )</td>
<td>0.033257</td>
<td>62.04171</td>
</tr>
<tr>
<td>( \hat{P}<em>{A</em>{2}(st)} )</td>
<td>0.032643</td>
<td>63.20877</td>
</tr>
<tr>
<td>( \hat{P}<em>{R</em>{1}(st)} )</td>
<td>0.004312</td>
<td>478.4485</td>
</tr>
<tr>
<td>( \hat{P}<em>{R</em>{2}(st)} )</td>
<td>0.004312</td>
<td>478.5419</td>
</tr>
<tr>
<td>( \hat{P}<em>{R</em>{3}(st)} )</td>
<td>0.004312</td>
<td>478.4688</td>
</tr>
<tr>
<td>( \hat{P}<em>{R</em>{4}(st)} )</td>
<td>0.004316</td>
<td>478.0895</td>
</tr>
<tr>
<td>( \hat{P}<em>{exp</em>{1}(st)} )</td>
<td>0.003271</td>
<td>630.7746</td>
</tr>
<tr>
<td>( \hat{P}<em>{exp</em>{2}(st)} )</td>
<td>0.003272</td>
<td>630.6883</td>
</tr>
<tr>
<td>( \hat{P}<em>{exp</em>{3}(st)} )</td>
<td>0.003272</td>
<td>630.5781</td>
</tr>
<tr>
<td>( \hat{P}<em>{exp</em>{4}(st)} )</td>
<td>0.003277</td>
<td>629.6699</td>
</tr>
</tbody>
</table>
We have computed the percent relative efficiencies (PREs) of $\hat{P}_{A, st}$ and $\hat{P}'_{A, st}$; with respect to usual unbiased estimator $t_{st} = p_{st}$ and displayed in Table 1.

From Table 1 it can be concluded that all proposed estimators $\hat{P}_{A, st}$ and $\hat{P}'_{A, st}$ are more efficient than the usual unbiased estimator [1] and Singh et al. (2010) estimators [2, 3] in the stratified random sampling.

REFERENCES

ESTIMATION OF FINITE POPULATION COEFFICIENT OF VARIATION IN STRATIFIED RANDOM SAMPLING

Aneel Ahmed\textsuperscript{1} and Javaid Shabbir\textsuperscript{2}
Department of Statistics, Quaid-e-Azam University, Islamabad, Pakistan
Email: sheikhs.88@hotmail.com; jsqau@yahoo.com

ABSTRACT

Using stratified random sampling scheme an estimator in estimating the finite population coefficient of variation (CV) when information on two auxiliary variables available is proposed. The properties of the proposed estimator are obtained. The special cases are also discussed by using different values of constant(s) and functions of known parameters. We consider the linear cost function to minimize the $MSE$ of the proposed estimator for a specified cost. The proposed estimator is more efficient than the usual sample CV, ratio estimator, Walsh (1970) estimator and Archana and Rao (2011) estimators. We use real data set for efficiency comparisons.

KEYWORDS

Auxiliary variable(s); CV; Mean square error ($MSE$); Efficiency.

1. INTRODUCTION

In the past several researchers have worked on the estimation of a finite population mean, population total and population variance but very little attention was given in the estimation of CV of the study variable ($y$) say $C_y$. The CV is widely used in agriculture, biology, economics and environmental sciences. Upadhyaya and Singh (1999), Khoshnevisan et al. (2007), Sisodia and Dwivedi (1981), etc suggested estimators in simple random sampling, Kadilar and Cingi (2003) and Shabbir and Gupta (2005) extended these estimators in stratified random sampling. Dalabehara and Sahoo (1999) developed a regression type estimator in stratified random sampling using two auxiliary variables. Singh and Vishwakarma (2008) suggested a family of estimators using transformation in stratified random sampling.


In Section 2, we introduce an estimator for finite population coefficient of variation by using two auxiliary variables and also discuss some existing estimators. Section 3 gives the efficiency comparison. Sections 4 and 5 give the empirical study and conclusion respectively.

Consider a finite population $U = \{1, 2, \ldots, N\}$ of size $N$, is stratified into $L$ strata. Let $y_{hi}$ and $(x_{hi}, z_{hi})$ be the observed values of the study variable ($y$) and the auxiliary
variables \((x, z)\) respectively for the \(i\)th unit \((i = 1, 2, \ldots, N)\) in the \(h\)th stratum consisting of \(N_h\) units \((h = 1, 2, \ldots, L)\) such that \(\sum_{h=1}^{L} N_h = N\). We draw a sample of size \(n_h\) from \(h\)th stratum by using simple random sample without replacement sampling scheme such that \(\sum_{h=1}^{L} n_h = n\). Let \(\bar{y}_h\) and \((\bar{x}_h, \bar{z}_h)\) be the sample means in \(h\)th stratum correspond to the population means \(\bar{Y}_h\) and \((\bar{X}_h, \bar{Z}_h)\) respectively.

Let \(s^2_{y_h} = \frac{\sum_{i=1}^{n_h} (y_{hi} - \bar{y}_h)^2}{n_h - 1}\), \(s^2_{x_h} = \frac{\sum_{i=1}^{n_h} (x_{hi} - \bar{x}_h)^2}{n_h - 1}\) and \(s^2_{z_h} = \frac{\sum_{i=1}^{n_h} (z_{hi} - \bar{z}_h)^2}{n_h - 1}\) be the sample variances in \(h\)th stratum correspond to the population variances \(S^2_{y_h} = \frac{\sum_{i=1}^{N_h} (y_{hi} - \bar{Y}_h)^2}{N_h - 1}\), \(S^2_{x_h} = \frac{\sum_{i=1}^{N_h} (x_{hi} - \bar{X}_h)^2}{N_h - 1}\) and \(S^2_{z_h} = \frac{\sum_{i=1}^{N_h} (z_{hi} - \bar{Z}_h)^2}{N_h - 1}\) respectively.

Let \(C_{y_h} = \frac{s^2_{y_h}}{\bar{y}_h}, C_{x_h} = \frac{s^2_{x_h}}{\bar{X}_h}\) and \(C_{z_h} = \frac{s^2_{z_h}}{\bar{Z}_h}\) be the coefficients of variation of \(y\), \(x\) and \(z\) respectively and \(W_h = \frac{N_h}{N}\) is the known stratum weight.

### 2. PROPOSED ESTIMATOR

Following Walsh, we propose a ratio type estimator for \(C_y\), given by

\[
\hat{C}_{yA_h} = \sum_{h=1}^{L} W_h \hat{C}_{yA_h},
\]

where,

\[
\hat{C}_{yA_h} = \hat{C}_{y_h} \left[ \frac{C^*_{x_h}}{\alpha_{4h} \hat{C}_{x_h} + (1 - \alpha_{4h}) C^*_{x_h}} \right] \left[ \frac{C^*_{z_h}}{\alpha_{2h} \hat{C}_{z_h} + (1 - \alpha_{2h}) C^*_{z_h}} \right],
\]

where \(C^*_{x_h} = a_h C_{x_h} + b_h, C^*_{z_h} = k_h C_{z_h} + l_h\), \(\hat{C}_{x_h} = a_h \hat{C}_{x_h} + b_h\) and \(\hat{C}_{z_h} = k_h \hat{C}_{z_h} + l_h\).

Here \(a_h \neq 0\), \(k_h \neq 0\), \(b_h\) and \(l_h\) are either real numbers or functions of known parameters which may be population mean \((\bar{U}_h)\), population coefficient of skewness \((\beta_1(u)_h)\), population coefficient of kurtosis \((\beta_2(u)_h)\), population correlation coefficient \((\rho_{yu})_h\), lower quartile \(Q_1(u)_h\) and upper quartile \(Q_3(u)_h\) for \(u = x, z\). Here \(\alpha_{4h} \in [0,1]\), \((i = 1,2)\) are constants whose values are to be determined. The aim of this paper is to
propose an efficient estimator of $C_y$ which should be better than the existing estimators. Also we generate some members of a proposed class of estimators.

To obtain the properties of $\hat{C}_{yA}$, we define the following relative error terms and their expectations.

Let $e_{0h} = \frac{y_h - \bar{Y}_h}{Y_h}$, $e_{1h} = \frac{x_h - \bar{X}_h}{X_h}$, $e_{2h} = \frac{z_h - \bar{Z}_h}{Z_h}$, $e_{3h} = \frac{s^2_{y_h} - s^2_{y_h}}{S^2_{y_h}}$, $e_{4h} = \frac{s^2_{x_h} - s^2_{x_h}}{S^2_{x_h}}$ and $e_{5h} = \frac{s^2_{z_h} - s^2_{z_h}}{S^2_{z_h}}$, such that $E(e_{ih}) = 0$, $i = 0-5$. Therefore

$$E(e_{0h}^2) = \theta_h C_{y_h}^2, \ E(e_{1h}^2) = \theta_h C_{x_h}^2, \ E(e_{2h}^2) = \theta_h C_{z_h}^2, \ E(e_{3h}^2) = \theta_h \left(\lambda_{400h} - 1\right),$$

$$E(e_{4h}^2) = \theta_h \left(\lambda_{040h} - 1\right), \ E(e_{5h}^2) = \theta_h \left(\lambda_{004h} - 1\right), \ E(e_{0h} e_{1h}) = \theta_h \rho_{xy_h} C_{y_h} C_{x_h},$$

$$E(e_{0h} e_{2h}) = \theta_h \rho_{xz_h} C_{y_h} C_{z_h}, \ E(e_{0h} e_{3h}) = \theta_h C_{y_h} \lambda_{300h}, \ E(e_{0h} e_{4h}) = \theta_h C_{y_h} \lambda_{120h},$$

$$E(e_{0h} e_{5h}) = \theta_h C_{y_h} \lambda_{102h}, \ E(e_{1h} e_{2h}) = \theta_h \rho_{xz_h} C_{x_h} C_{z_h}, \ E(e_{1h} e_{3h}) = \theta_h C_{x_h} \lambda_{210h},$$

$$E(e_{1h} e_{4h}) = \theta_h C_{x_h} \lambda_{102h}, \ E(e_{1h} e_{5h}) = \theta_h C_{x_h} \lambda_{030h}, \ E(e_{2h} e_{3h}) = \theta_h C_{z_h} \lambda_{012h},$$

$$E(e_{2h} e_{4h}) = \theta_h C_{z_h} \lambda_{021h}, \ E(e_{2h} e_{5h}) = \theta_h C_{z_h} \lambda_{201h}, \ E(e_{3h} e_{4h}) = \theta_h C_{z_h} \lambda_{003h}, \ E(e_{3h} e_{5h}) = \theta_h \left(\lambda_{220h} - 1\right),$$

$$E(e_{4h} e_{5h}) = \theta_h \left(\lambda_{202h} - 1\right), \ E(e_{4h} e_{5h}) = \theta_h \left(\lambda_{022h} - 1\right),$$

where $\theta_h = \left(1 - \frac{f_h}{n_h}\right), \ f_h = \frac{n_h}{N_h}, \ \lambda_{rst_h} = \frac{\mu_{rst_h}}{\mu_{200h} r/2 \mu_{020h} s/2 \mu_{002h} t/2}$ and

$$\mu_{rst_h} = \frac{\sum_{i=1}^{N_h} \left(y_{hi} - \bar{Y}_h\right) r \left(x_{hi} - \bar{X}_h\right) s \left(z_{hi} - \bar{Z}_h\right) t}{N_h - 1}.$$
\[
\hat{C}_{yA_h} = \frac{\left(\frac{s_{y_h}}{y_h}\right)^{1/2}}{y_h} \left[ \frac{a_h C_{y_h} + b_h}{\alpha_{1_h} \left\{ a_h \left(\frac{s_{x_h}}{x_h}\right)^{1/2} + b_h \right\} + \left(1 - \alpha_{1_h}\right) \left\{ a_h C_{x_h} + b_h \right\}} \right]
\]

\[
= \frac{k_h C_{z_h} + l_h}{\alpha_{2_h} \left\{ k_h \left(\frac{s_{z_h}}{z_h}\right)^{1/2} + l_h \right\} + \left(1 - \alpha_{2_h}\right) \left\{ k_h C_{z_h} + l_h \right\}}
\]

In terms of \( e ' s \), we have

\[
\hat{C}_{yA_h} = C_{y_h} \left( \frac{1+e_{3_h}}{1+e_{0_h}} \right)^{1/2} \left[ \frac{a_h C_{y_h} + b_h}{\alpha_{1_h} \left\{ a_h \left(1+e_{4_h}\right)^{1/2} + b_h \right\} + \left(1 - \alpha_{1_h}\right) \left\{ a_h C_{x_h} + b_h \right\}} \right]
\]

\[
= \frac{k_h C_{z_h} + l_h}{\alpha_{2_h} \left\{ k_h \left(1+e_{5_h}\right)^{1/2} + l_h \right\} + \left(1 - \alpha_{2_h}\right) \left\{ k_h C_{z_h} + l_h \right\}}
\]

Expanding the right hand side of above equation, neglecting terms of \( e ' s \) having power greater than two, we have

\[
\left(\hat{C}_{yA_h} - C_{y_h}\right) \approx C_{y_h} \left[ \left\{ -e_{0_h} + e_{0_h}^2 + \frac{1}{2} e_{3_h}^2 - \frac{1}{2} e_{0_h} e_{3_h} - \frac{1}{8} e_{3_h}^2 \right\} \right]
\]

\[
- \alpha_{2_h} P_h \left\{ -e_{2_h} + e_{2_h}^2 + \frac{1}{2} e_{5_h}^2 - \frac{1}{2} e_{2_h} e_{5_h} - \frac{1}{8} e_{5_h}^2 + e_{0_h} e_{2_h} \right\} \]

\[
- \frac{1}{2} e_{0_h} e_{3_h} - \frac{1}{2} e_{2_h} e_{3_h} + \frac{1}{4} e_{3_h} e_{5_h} \right\}
\]
\[ -\alpha_{1h} T_h \left( -e_{1h} + e_{1h}^2 + \frac{1}{2} e_{1h}^4 - \frac{1}{2} e_{1h} e_{4h} - \frac{1}{8} e_{4h}^2 \right) \]
\[ + e_{0h} e_{1h} - \frac{1}{2} e_{0h} e_{4h} - \frac{1}{2} e_{1h} e_{3h} + \frac{1}{4} e_{3h} e_{4h} \right) + \alpha_{2h}^2 P_h \left( \frac{1}{4} e_{5h}^2 + e_{2h}^2 - e_{2h} e_{5h} \right) \]
\[ + \alpha_{1h} \alpha_{2h} P_h T_h \left( e_{1h} e_{2h} - \frac{1}{2} e_{1h} e_{5h} - \frac{1}{2} e_{2h} e_{4h} + \frac{1}{4} e_{4h} e_{5h} \right) \]
\[ + \alpha_{1h}^2 T_h^2 \left( \frac{1}{4} e_{4h}^2 + e_{1h}^2 - e_{1h} e_{4h} \right) \],

where
\[ T_h = \frac{d_h C_{xh}}{a_h C_{xh} + b_h} \quad \text{and} \quad P_h = \frac{k_h C_{z_h}}{k_h C_{z_h} + l_h}. \]

Taking expectation on both sides of (3), we get
\[ \text{Bias}\left( \hat{C}_{yA} \right) \approx C_{y} \theta_h \left[ G_h - \alpha_{1h} T_h \left( I_h + E_h \right) - \alpha_{2h} P_h \left( H_h + D_h \right) \right. \]
\[ \left. + \alpha_{1h}^2 T_h^2 C_h + \alpha_{2h}^2 P_h^2 B_h + \alpha_{1h} \alpha_{2h} P_h T_h F_h \right] \]

Substituting the value of \( \text{Bias}\left( \hat{C}_{yA} \right) \) in (1), we get the bias of \( \hat{C}_{yA} \), given by
\[ \text{Bias}\left( \hat{C}_{yAS} \right) \approx \sum_{h=1}^{L} W_h C_{y} \theta_h \left[ G_h - \alpha_{1h} T_h \left( I_h + E_h \right) - \alpha_{2h} P_h \left( H_h + D_h \right) \right. \]
\[ \left. + \alpha_{1h}^2 T_h^2 C_h + \alpha_{2h}^2 P_h^2 B_h + \alpha_{1h} \alpha_{2h} P_h T_h F_h \right] \]

where
\[ B_h = (1/4)\left( \lambda_{004h} - 1 \right) + C_{z_h}^2 - C_{z_h} \lambda_{003h} \]
\[ C_h = (1/4)\left( \lambda_{040h} - 1 \right) + C_{z_h}^2 - C_{z_h} \lambda_{030h} \]
\[ D_h = (1/4)\left( \lambda_{202h} - 1 \right) - (1/2) C_{y} \lambda_{102h} - (1/2) C_{z_h} \lambda_{021h} + \rho_{y_{0z}} C_{y} C_{z_h} \]
\[ E_h = (1/4)\left( \lambda_{220h} - 1 \right) - (1/2) C_{y} \lambda_{120h} - (1/2) C_{z_h} \lambda_{210h} + \rho_{y_{0z}} C_{y} C_{z_h} \]
\[ F_h = (1/4)\left( \lambda_{022h} - 1 \right) - (1/2) C_{y} \lambda_{012h} - (1/2) C_{z_h} \lambda_{021h} + \rho_{y_{0z}} C_{y} C_{z_h} \]
\[ G_h = C_{y}^2 - (1/2) C_{y} \lambda_{300h} - (1/8) \left( \lambda_{400h} - 1 \right) \]
\[ H_h = C_{z_h}^2 - (1/2) C_{z_h} \lambda_{303h} - (1/8) \left( \lambda_{403h} - 1 \right) \]
\[ I_h = C_{x}^2 - (1/2) C_{x} \lambda_{030h} - (1/8) \left( \lambda_{040h} - 1 \right). \]

Squaring both sides of (3), and keeping terms upto order 2, we have,
\[
\left( \hat{C}_{yA_h} - C_{y_h} \right)^2 \approx C_{y_h}^2 \left[ \frac{1}{4} e_3^2 + e_0^2 - e_0 e_3 \right] + \alpha_2^2 \frac{P_h}{2} \left( \frac{1}{4} e_3^2 + e_0^2 - e_2 e_5 \right)
\]
\[
+ \alpha_1^2 \frac{T_h}{2} \left( \frac{1}{4} e_3^2 + e_1 e_4 \right) - 2T_h \alpha_1 \left( \frac{1}{4} e_3^2 e_4 - \frac{1}{2} e_1 e_3 - \frac{1}{2} e_0 e_4 + e_0 e_1 \right)
\]
\[
+ 2P_h T_h \alpha_1 \alpha_2 \left( \frac{1}{4} e_4 e_5 - \frac{1}{2} e_1 e_3 \right) + 2P_h T_h \alpha_1 \alpha_2 \left( e_1 e_2 - \frac{1}{2} e_2 e_4 \right)
\]
\[
- 2P_h \alpha_2 \left( \frac{1}{4} e_3 e_5 - \frac{1}{2} e_2 e_3 - \frac{1}{2} e_0 e_5 + e_0 e_2 \right) \].
\]

Taking the expectation on both sides of (4), we get

\[
MSE(\hat{C}_{yA_h}) \approx C_{y_h}^2 \left[ A_h + \alpha_1^2 \frac{T_h}{2} C_h + \alpha_2^2 \frac{P_h}{2} B_h - 2\alpha_1 T_h E_h \right]
\]
\[
- 2\alpha_2 P_h D_h + 2\alpha_1 \alpha_2 P_h T_h F_h \],
\]

where

\[
A_h = \left( \frac{1}{4} \right) (\lambda_{400h} - 1) + C_{y_h}^2 - C_{y_h} \lambda_{300h}.
\]

Substituting the value of \( MSE(\hat{C}_{yA_h}) \) in (1), we get the \( MSE \) of \( \hat{C}_{yA_S} \), given by

\[
MSE(\hat{C}_{yA_S}) \approx \sum_{h=1}^{L} W_h^2 C_{y_h}^2 \left[ A_h + \alpha_1^2 \frac{T_h}{2} C_h + \alpha_2^2 \frac{P_h}{2} B_h - 2\alpha_1 T_h E_h \right]
\]
\[
- 2\alpha_2 P_h D_h + 2\alpha_1 \alpha_2 P_h T_h F_h \],
\]

Using (5), the optimum values of \( \alpha_{1_h} \) and \( \alpha_{2_h} \), are given by

\[
\alpha_{1_h}(opt) = \frac{B_h E_h - D_h F_h}{T_h \left( B_h C_h - F_h^2 \right)} \quad \text{and} \quad \alpha_{2_h}(opt) = \frac{C_h D_h - E_h F_h}{P_h \left( B_h C_h - F_h^2 \right)}.
\]

Substituting the optimum values of \( \alpha_{i_h} \) \( (i = 1, 2) \) in (5), we get the minimum \( MSE \) of \( \hat{C}_{yA_S} \), given by

\[
MSE(\hat{C}_{yA_S})_{\text{min}} \approx \sum_{h=1}^{L} W_h^2 C_{y_h}^2 \left[ A_h - \frac{C_h D_h^2 + B_h E_h^2 - 2D_h E_h F_h}{B_h C_h - F_h^2} \right].
\]

3. SPECIAL CASES

(i) If both \( \alpha_{1_h} \) and \( \alpha_{2_h} \) are equal to zero in (2), then \( \hat{C}_{yA_h} \) reduces to

\[
\hat{C}_{yA_{h(i)}} = \hat{C}_{y}.
\]

The \( Bias \) and \( MSE \) of \( \hat{C}_{yA_{h(i)}} \), to first degree approximation, are given by
Ahmed and Shabbir

\[ \text{Bias} \left( \hat{C}_{yA}^{(1)} \right) \approx C_{y_h} \theta_h G_h \] (8)

and

\[ \text{MSE} \left( \hat{C}_{yA}^{(1)} \right) \approx C_{y_h}^2 \theta_h A_h . \] (9)

As,

\[ \hat{C}_{yA}^{(1)} = \sum_{h=1}^{L} W_h \hat{C}_{yA}^{(1)} . \] (10)

Substituting Bias and MSE of \( \hat{C}_{yA}^{(1)} \) in (10), we get the Bias and MSE of \( \hat{C}_{yA}^{(1)} \), to first degree approximation, are given by

\[ \text{Bias} \left( \hat{C}_{yA}^{(1)} \right) \approx \sum_{h=1}^{L} W_h C_{y_h} \theta_h G_h \] (11)

and

\[ \text{MSE} \left( \hat{C}_{yA}^{(1)} \right) \approx \sum_{h=1}^{L} W_h^2 C_{y_h}^2 \theta_h A_h \] (12)

(ii) If \( \alpha_{y_h} = 0 \) in (2), then \( \hat{C}_{yA}^{(1)} \) reduces to Walsh (1970) estimator as

\[ \hat{C}_{yA}^{(2)} = \hat{C}_{y_h} \left[ \frac{k_h C_{z_h} + l_h}{\alpha_{2_h} \left\{ k_h \hat{C}_{z_h} + l_h \right\} + \left( 1 - \alpha_{2_h} \right) \left\{ k_h C_{z_h} + l_h \right\} } \right] = \hat{C}_{yW} . \] (13)

The Bias and MSE of \( \hat{C}_{yA}^{(2)} \), to first degree approximation, are given by

\[ \text{Bias} \left( \hat{C}_{yA}^{(2)} \right) \approx C_{y_h} \theta_h \left[ G_h - \alpha_{2_h} P_h \left( H_h + D_h \right) + \alpha_{2_h}^2 P \ B \right] \] (14)

and

\[ \text{MSE} \left( \hat{C}_{yA}^{(2)} \right) \approx C_{y_h}^2 \theta_h \left[ A_h + \alpha_{2_h}^2 P_h^2 B_h - 2 \alpha_{2_h} P_h D_h \right] . \] (15)

As,

\[ \hat{C}_{yA}^{(2)} = \sum_{h=1}^{L} W_h \hat{C}_{yA}^{(2)} \] (16)

Substituting MSE of \( \hat{C}_{yA}^{(2)} \) in (16), we get the MSE of \( \hat{C}_{yA}^{(2)} \), to first degree approximation, is given by

\[ \text{MSE} \left( \hat{C}_{yA}^{(2)} \right) \approx \sum_{h=1}^{L} W_h^2 C_{y_h}^2 \theta_h \left[ A_h + \alpha_{2_h}^2 P_h^2 B_h - 2 \alpha_{2_h} P_h D_h \right] . \] (17)

Using (17), the optimum value of \( \alpha_{2_h} \), is given by

\[ \alpha_{2_h (opt)} = \frac{D_h}{B_h P_h} . \]

Substituting the optimum value of \( \alpha_{2_h} \) in (17), we get the minimum MSE of \( \hat{C}_{yA}^{(2)} \), given by
Estimation of Finite Population Coefficient of Variation

\[ \text{MSE} \left( \hat{C}_{yA}^{(2)} \right)_{\min} \approx \sum_{h=1}^{L} W_h^2 C_{y_h}^2 \theta_h A_h \left[ 1 - \frac{D_h^2}{A_h B_h} \right] = \text{MSE} \left( \hat{C}_{yW} \right)_{\min} \quad (18) \]

(iii) If \( \alpha_{2h} = 0 \) in (2), then \( \hat{C}_{yA_h} \) reduces to another Walsh (1970) estimator as

\[ \hat{C}_{yA_h}^{(3)} = \hat{C}_{y_h} \left[ \frac{a_h C_{x_h} + b_h}{\alpha_{1h} \left( a_h \hat{C}_{x_h} + b_h \right) + \left( 1 - \alpha_{1h} \right) \left( a_h C_{x_h} + b_h \right)} \right] = \hat{C}_{yW_h}. \quad (19) \]

The Bias and MSE of \( \hat{C}_{yA_h}^{(3)} \), to first degree approximation, are given by

\[ \text{Bias} \left( \hat{C}_{yA_h}^{(3)} \right) \approx C_{y_h} \theta_h \left[ G_h - \alpha_{1h} T_h \left( I_h + E_h \right) + \alpha_{1h}^2 T_h^2 C_h \right] \quad (20) \]

and

\[ \text{MSE} \left( \hat{C}_{yA_h}^{(3)} \right) \approx C_{y_h}^2 \theta_h \left[ A_h + \alpha_{1h}^2 T_h^2 C_h - 2\alpha_{1h} T_h E_h \right]. \quad (21) \]

As,

\[ \hat{C}_{yA_h}^{(3)} = \sum_{h=1}^{L} W_h \hat{C}_{yA_h}^{(3)} \quad (22) \]

Substituting MSE of \( \hat{C}_{yA_h}^{(3)} \) in (22), we get the MSE of \( \hat{C}_{yA}^{(3)} \), to first degree approximation, is given by

\[ \text{MSE} \left( \hat{C}_{yA_h}^{(3)} \right) \approx \sum_{h=1}^{L} W_h^2 C_{y_h}^2 \theta_h \left[ A_h + \alpha_{1h}^2 T_h^2 C_h - 2\alpha_{1h} T_h E_h \right]. \quad (23) \]

Using (23), the optimum value of \( \alpha_{1h} \), is given by

\[ \alpha_{1h,(opt)} = \frac{E_h}{C_h T_h}. \]

Substituting the optimum value of \( \alpha_{1h} \) in (23), we get the minimum MSE of \( \hat{C}_{yA}^{(3)} \), given by

\[ \text{MSE} \left( \hat{C}_{yA}^{(3)} \right)_{\min} \approx \sum_{h=1}^{L} W_h^2 C_{y_h}^2 \theta_h A_h \left[ 1 - \frac{E_h^2}{A_h C_h} \right] = \text{MSE} \left( \hat{C}_{yW_h} \right)_{\min} \quad (24) \]

(iv) If \( \alpha_{1h} = 1 \) in (2), then \( \hat{C}_{yA_h} \) reduces to

\[ \hat{C}_{yA_h}^{(4)} = \hat{C}_{y_h} \left[ \frac{a_h C_{x_h} + b_h}{a_h \hat{C}_{x_h} + b_h} \right] \left[ \frac{k_h C_{z_h} + l_h}{\alpha_{2h} \left( k_h \hat{C}_{z_h} + l_h \right) + \left( 1 - \alpha_{2h} \right) \left( k_h C_{z_h} + l_h \right)} \right]. \quad (25) \]

The Bias and MSE of \( \hat{C}_{yA_h}^{(4)} \), to first degree approximation, are given by
\[ Bias\left( \hat{C}_{yA}^{(4)} \right) \approx C_{yA} \theta \begin{bmatrix} G_h - \alpha_2 P_h (H_h + D_h) + \alpha_2^2 P_h^2 B_h \\ -T_h (I_h + E_h) + \alpha_2 T_h P_h F_h + T_h^2 C_h \end{bmatrix} \] (26)

and

\[ MSE\left( \hat{C}_{yA}^{(4)} \right) \approx C_{yA}^2 \theta \begin{bmatrix} A_h + \alpha_2^2 P_h^2 B_h + T_h^2 C_h - 2\alpha_2 P_h D_h \\ -2T_h E_h + 2\alpha_2 P_h T_h F_h \end{bmatrix} \] (27)

As, \[ \hat{C}_{yA}^{(4)} = \sum_{h=1}^{L} W_h \hat{C}_{yA}^{(4)} \] (28)

Substituting \( MSE \) of \( \hat{C}_{yA}^{(4)} \) in (28), we get the \( MSE \) of \( \hat{C}_{yA}^{(5)} \), to first degree approximation, is given by

\[ MSE\left( \hat{C}_{yA}^{(5)} \right) \approx \sum_{h=1}^{L} W_h^2 C_{yA}^2 \theta \begin{bmatrix} A_h + \alpha_2^2 P_h^2 B_h + T_h^2 C_h - 2\alpha_2 P_h D_h \\ -2T_h E_h + 2\alpha_2 P_h T_h F_h \end{bmatrix} \] (29)

Using (29), the optimum value of \( \alpha_2 \), is given by

\[ \alpha_2_{(opt)} = \frac{D_h - F_h T_h}{B_h P_h} \].

Substituting the optimum value of \( \alpha_2 \) in (29), we get the minimum \( MSE \) of \( \hat{C}_{yA}^{(5)} \), given by

\[ MSE\left( \hat{C}_{yA}^{(5)} \right)_{\text{min}} \approx \sum_{h=1}^{L} W_h^2 C_{yA}^2 \theta \begin{bmatrix} 1 - \left( \frac{D_h - F_h T_h}{B_h P_h} \right)^2 - B_h C_h T_h^2 + 2B_h E_h T_h \end{bmatrix} \frac{A_h B_h}{A_h B_h} \] (30)

(v) If \( \alpha_2 = 1 \) in (2), then \( \hat{C}_{yA} \) reduces to

\[ \hat{C}_{yA}^{(5)} = \hat{C}_{yA} \left[ \frac{a_h C_{yA} + b_h}{\alpha_1_h \left( a_h \hat{C}_{yA} + b_h \right) + \left( 1 - \alpha_1_h \right) \left( a_h C_{yA} + b_h \right)} \right] \left[ \frac{k_h C_{yA} + l_h}{k_h \hat{C}_{yA} + l_h} \right]. \] (31)

The \( Bias \) and \( MSE \) of \( \hat{C}_{yA}^{(5)} \), to first degree approximation, are given by

\[ Bias\left( \hat{C}_{yA}^{(5)} \right) \approx C_{yA} \theta \begin{bmatrix} G_h - P_h (H_h + D_h) + P_h^2 B_h - \alpha_1 T_h (I_h + E_h) \\ + \alpha_1 T_h P_h F_h + \alpha_1^2 T_h^2 C_h \end{bmatrix} \] (32)

and

\[ MSE\left( \hat{C}_{yA}^{(5)} \right) \approx C_{yA}^2 \theta \begin{bmatrix} A_h + P_h^2 B_h + \alpha_1^2 T_h^2 C_h - 2P_h D_h - 2\alpha_1 T_h E_h + 2\alpha_1 P_h T_h F_h \end{bmatrix}. \] (33)

As, \[ \hat{C}_{yA}^{(5)} = \sum_{h=1}^{L} W_h \hat{C}_{yA}^{(5)} \] (34)
Substituting \( MSE \) of \( \hat{C}_{yA_h}^{(5)} \) in (34), we get the \( MSE \) of \( \hat{C}_{yA}^{(5)} \), to first degree approximation, is given by
\[
MSE(\hat{C}_{yA}^{(5)}) \cong \frac{L}{h=1} W_h^2 C_{y_h}^2 \theta_h \left[ A_h + P_h^2 B_h + \alpha_{1h}^2 T_h^2 C_h - 2P_h D_h - 2\alpha_{1h} T_h E_h + 2\alpha_{1h} P_h T_h F_h \right]
\] (35)

Using (35), the optimum value of \( \alpha_{1h} \), is given by
\[
\alpha_{1h,(opt)} = \frac{E_h - F_h P_h}{C_h T_h}.
\]

Substituting the optimum value of \( \alpha_{1h} \) in (35), we get the minimum \( MSE \) of \( \hat{C}_{yA}^{(5)} \), given by
\[
MSE(\hat{C}_{yA}^{(5)})_{min} \cong \frac{L}{h=1} W_h^2 C_{y_h}^2 \theta_h A_h \left[ 1 - \frac{(E_h - F_h P_h)^2 - B_h C_h P_h^2 + 2C_h D_h P_h}{A_h C_h} \right].
\] (36)

(vi) If both \( \alpha_{1h} \) and \( \alpha_{2h} \) are equal to one i.e \( \alpha_{1h} = \alpha_{2h} = 1 \) in (2), then \( \hat{C}_{yA} \) reduces to
\[
\hat{C}_{yA}^{(6)} = \hat{C}_{y_h} \left[ \frac{a_h C_{y_h} + b_h}{a_h \hat{C}_{y_h} + b_h} \right] \left[ \frac{k_h C_{z_h} + l_h}{k_h \hat{C}_{z_h} + l_h} \right].
\] (37)

The \( Bias \) and \( MSE \) of \( \hat{C}_{yA}^{(6)} \), to first degree approximation, are given by
\[
Bias(\hat{C}_{yA}^{(6)}) \cong C_{y_h} \theta_h G_h \left[ 1 - \frac{P_h (H_h + D_h) + T_h (I_h + E_h) - B_h P_h^2 - F_h P_h T_h - C_h T_h^2}{G_h} \right]
\] (38)

and
\[
MSE(\hat{C}_{yA}^{(6)}) \cong C_{y_h}^2 \theta_h A_h \left[ 1 - \frac{2D_h P_h + 2E_h T_h - 2F_h P_h T_h - B_h P_h^2 - C_h T_h^2}{A_h} \right].
\] (39)

As, \( \hat{C}_{yA}^{(6)} = \sum_{h=1}^{L} W_h \hat{C}_{yA}^{(6)} \)
(40)

Substituting \( Bias \) and \( MSE \) of \( \hat{C}_{yA}^{(6)} \) in (40), we get the \( Bias \) and \( MSE \) of \( \hat{C}_{yA}^{(6)} \), to first degree approximation, are given by
\[
Bias(\hat{C}_{yA}^{(6)}) \cong \sum_{h=1}^{L} W_h C_{y_h} \theta_h G_h \left[ 1 - \frac{P_h (H_h + D_h) + T_h (I_h + E_h) - B_h P_h^2 - F_h P_h T_h - C_h T_h^2}{G_h} \right]
\] (41)

and
\[
MSE(\hat{C}_{yA}^{(6)}) \cong \sum_{h=1}^{L} W_h^2 C_{y_h}^2 \theta_h A_h \left[ 1 - \frac{2D_h P_h + 2E_h T_h - 2F_h P_h T_h - B_h P_h^2 - C_h T_h^2}{A_h} \right].
\] (42)
(vii) If \( (\alpha_{1h}, \alpha_{2h}) = (1, 0) \) and \( (a_h, b_h) = (1, 0) \) in (2), then \( \hat{C}_{yA_h} \) reduces to
\[
\hat{C}_{yR_h} = \hat{C}_{y_h} \left[ \frac{C_{y_h}}{C_{z_h}} \right].
\] (43)

The Bias and MSE of \( \hat{C}_{yR_h} \), to first degree approximation, are given by
\[
Bias\left( \hat{C}_{yR_h} \right) \approx C_{y_h} \theta_h \left[ C_h - E_h + G_h - I_h \right]
\] (44) and
\[
MSE\left( \hat{C}_{yR_h} \right) \approx C_{y_h}^2 \theta_h \left[ A_h + C_h - 2E_h \right].
\] (45)

As,
\[
\hat{C}_{yR_S} = \sum_{h=1}^{L} W_h \hat{C}_{yR_h}
\] (46)

Substituting Bias and MSE of \( \hat{C}_{yR_h} \) in (46), we get the Bias and MSE of \( \hat{C}_{yR_S} \), to first degree approximation, are given by
\[
Bias\left( \hat{C}_{yR_S} \right) \approx \sum_{h=1}^{L} W_h C_{y_h} \theta_h \left[ C_h - E_h + G_h - I_h \right]
\] (47) and
\[
MSE\left( \hat{C}_{yR_S} \right) \approx \sum_{h=1}^{L} W_h^2 C_{y_h}^2 \theta_h \left[ A_h + C_h - 2E_h \right]
\] (48)

(viii) If \( (\alpha_{1h}, \alpha_{2h}) = (0, 1) \) and \( (k_h, l_h) = (1, 0) \) in (2), then \( \hat{C}_{yA_h} \) reduces to
\[
\hat{C}_{yS} = \hat{C}_{y_h} \left[ \frac{C_{z_h}}{C_{z_h}} \right].
\] (49)

The Bias and MSE of \( \hat{C}_{yS} \), to first degree approximation, are given by
\[
Bias\left( \hat{C}_{yS} \right) \approx C_{y_h} \theta_h \left[ B_h - D_h + G_h - H_h \right]
\] (50) and
\[
MSE\left( \hat{C}_{yS} \right) \approx C_{y_h}^2 \theta_h \left[ A_h + B_h - 2D_h \right].
\] (51)

As,
\[
\hat{C}_{yS} = \sum_{h=1}^{L} W_h \hat{C}_{yS}
\] (52)

Substituting Bias and MSE of \( \hat{C}_{yS} \) in (52), we get the Bias and MSE of \( \hat{C}_{yS} \), to first degree approximation, are given by
\[
Bias\left( \hat{C}_{yS} \right) \approx \sum_{h=1}^{L} W_h C_{y_h} \theta_h \left[ B_h - D_h + G_h - H_h \right]
\] (53)
and
\[
MSE\left(\hat{C}_{yRAS}\right) \approx \sum_{h=1}^{L} W_h^2 C_{y_h}^2 \theta_h \left[A_h + B_h - 2D_h\right]
\]  \hspace{1cm} (54)

(ix) Archana and Rao (2011) ratio estimator is given by
\[
\hat{C}_{yRA_h}^{(1)} = \hat{C}_{y_h} \left[\frac{\bar{X}_h}{x_h}\right].
\]  \hspace{1cm} (55)

The Bias and MSE of \(\hat{C}_{yRA_h}^{(1)}\), to first degree approximation, are given by
\[
Bias\left(\hat{C}_{yRA_h}^{(1)}\right) \approx C_{y_h} \theta_h G_h \left[1 - \frac{J_h}{G_h}\right],
\]  \hspace{1cm} (56)
where \(J_h = \left(1/2\right) C_{x_h} \lambda_{210_h} - C_{x_h}^2 - \rho_{yx_h} C_{y_h} C_{x_h}\)
and
\[
MSE\left(\hat{C}_{yRA_h}^{(1)}\right) \approx C_{y_h}^2 \theta_h A_h \left[1 - \frac{M_h}{A_h}\right],
\]  \hspace{1cm} (57)
where \(M_h = C_{x_h} \lambda_{210_h} - C_{x_h}^2 - 2\rho_{yx_h} C_{y_h} C_{x_h}\).

As, \(\hat{C}_{yRA_S} = \sum_{h=1}^{L} W_h \hat{C}_{yRA_h}^{(1)}\)  \hspace{1cm} (58)
Substituting Bias and MSE of \(\hat{C}_{yRA_h}^{(1)}\) in (58), we get the Bias and MSE of \(\hat{C}_{yRA_S}^{(1)}\), to first degree approximation, are given by
\[
Bias\left(\hat{C}_{yRA_S}^{(1)}\right) \approx \sum_{h=1}^{L} W_h C_{y_h} \theta_h G_h \left[1 - \frac{J_h}{G_h}\right]
\]  \hspace{1cm} (59)
and
\[
MSE\left(\hat{C}_{yRA_S}^{(1)}\right) \approx \sum_{h=1}^{L} W_h^2 C_{y_h}^2 \theta_h A_h \left[1 - \frac{M_h}{A_h}\right]
\]  \hspace{1cm} (60)

(x) Yet another Archana and Rao (2011) ratio estimator is given by
\[
\hat{C}_{yRA_h}^{(2)} = C_{y_h} \left[\frac{S_y^2}{S_{x_h}^2}\right].
\]  \hspace{1cm} (61)

The Bias and MSE of \(\hat{C}_{yRA_h}^{(2)}\), to first degree approximation, are given by
\[
Bias\left(\hat{C}_{yRA_h}^{(2)}\right) \approx C_{y_h} \theta_h G_h \left[1 - \frac{N_h}{G_h}\right],
\]  \hspace{1cm} (62)
where \(N_h = \left(1/2\right)\left(\lambda_{220_h} - 1\right) - C_{x_h} \lambda_{120_h} - \left(\lambda_{040_h} - 1\right)\)
and
\[
MSE \left( \hat{C}_{yRA_h}^{(2)} \right) \approx C_{yh}^2 \theta_h A_h \left[ 1 - \frac{O_h}{A_h} \right],
\]
where \( O_h = \lambda_{220h} - 2C_{yh} \lambda_{120h} - \lambda_{040h} \).

As, \( \hat{C}_{yRA}^{(2)} = \sum_{h=1}^{L} W_h \hat{C}_{yRA_h}^{(2)} \) (64)

Substituting Bias and MSE of \( \hat{C}_{yRA_h}^{(2)} \) in (64), we get the Bias and MSE of \( \hat{C}_{yRA}^{(2)} \), to first degree approximation, are given by

\[
Bias \left( \hat{C}_{yRA}^{(2)} \right) \approx \sum_{h=1}^{L} W_h C_{yh}^2 \theta_h G_h \left[ 1 - \frac{N_h}{G_h} \right]
\]

and

\[
MSE \left( \hat{C}_{yRA}^{(2)} \right) \approx \sum_{h=1}^{L} W_h^2 C_{yh}^2 \theta_h A_p \left[ 1 - \frac{O_h}{A_h} \right]
\]

4. LINEAR COST FUNCTION

To minimize the \( MSE(\hat{C}_{yS}^{(1)}) \), \( MSE(\hat{C}_{yq}) \), where \( q = R_1S \), \( R_2S \), \( RA_1 \), \( RA_2 \), \( MSE(\hat{C}_{yp})_{\text{min}} \) where \( p = A_S \), \( W_1S \), \( W_2S \) we consider the linear cost function.

i.e.,

\[
C' = C_0' + \sum_{h=1}^{L} c'_h n_h
\]

where \( C' \) is the total cost, \( C_0' \) is the fixed cost, \( c'_h \) is the per unit cost of measurement in the \( h \) th stratum and \( n_h \) is the sample size from the \( h \) th stratum to minimized the above mentioned MSE for a specified cost.

Using the Lagrange’s method of multipliers we find the integral value of \( n_h \) through rounding up.

We define Lagrangian function as,

\[
L(n_h, \lambda) = \sum_{h=1}^{L} W_h C_{yh}^2 \theta_h \left[ A_h - \frac{C_h D_h^2 + B_h E_h^2 - 2D_h E_h F_h}{B_h C_h - F_h^2} \right] + \lambda \left( C_0' + \sum_{h=1}^{L} c'_h n_h - C'ight)
\]

where \( \lambda \) is the Lagrange Multiplier.

Taking the partial derivative of (67) w.r.t. \( n_h \); \( h = 1,2,\ldots,L \) and \( \lambda \) and equating to zero, we get \( L+1 \) equations as

\[
\frac{\partial L(n_h, \lambda)}{\partial n_h} = - \frac{W_h^2 C_{yh}^2}{n_h^2} \left( A_h - \frac{C_h D_h^2 + B_h E_h^2 - 2D_h E_h F_h}{B_h C_h - F_h^2} \right) + \lambda c'_h = 0
\]

and
\[
\frac{\partial L(n_h, \lambda)}{\partial \lambda} = C'_0 + \sum_{h=1}^{L} c'_h n_h - C' = 0
\]  

(69)

From (68), we get \( n_h \) as:
\[
n_h = \frac{W_h C_{\gamma h} \sqrt{A_h - \frac{C_h D_h^2 + B_h E_h^2 - 2D_h E_h F_h}{B_h C_h - F_h^2}}}{\sqrt{\lambda c_h}}
\]  

(70)

Substituting \( n_h \) from (70) in (69), we get
\[
\frac{1}{\sqrt{\lambda}} = \left( C' - C'_0 \right)
\]

(71)

Now substituting \( \frac{1}{\sqrt{\lambda}} \) from (71) in (70), we get
\[
n_h = \frac{W_h C_{\gamma h} \sqrt{A_h - \frac{C_h D_h^2 + B_h E_h^2 - 2D_h E_h F_h}{B_h C_h - F_h^2}}}{\sqrt{\lambda c_h}}
\]

(72)

As we see that from (72), \( n_h \) increases as fixed budget \( \left( C' - C'_0 \right) \) increases. So fixed budget would be minimum.

5. EFFICIENCY COMPARISON

We compare the proposed estimator with other existing estimators. By (6) and (12),
\[
MSE\left( \hat{\gamma}_{yLS} \right)_{\min} < MSE\left( \hat{\gamma}_{yLS}^{(i)} \right) \quad \text{if} \quad \sum_{h=1}^{L} W_h^2 C_{\gamma h}^2 \theta_h \left( \frac{C_h D_h^2 + B_h E_h^2 - 2D_h E_h F_h}{B_h C_h - F_h^2} \right) > 0 .
\]

(i) By (6) and (18),
\[
MSE\left( \hat{\gamma}_{yLS} \right)_{\min} < MSE\left( \hat{\gamma}_{yWS} \right)_{\min} \quad \text{if} \quad \sum_{h=1}^{L} W_h^2 C_{\gamma h}^2 \theta_h \left( \frac{C_h D_h^2 + B_h E_h^2 - 2D_h E_h F_h}{B_h C_h - F_h^2} - \frac{D_h^2}{B_h} \right) > 0 .
\]
(ii) By (6) and (24),
\[ \text{MSE}\left(\hat{C}_{yA_S}\right)_{\min} < \text{MSE}\left(\hat{C}_{yW_2S}\right)_{\min} \text{ if } \]
\[ \sum_{h=1}^{L} W_h^2 C_{y_h}^2 \theta_h \left( \frac{C_h D_h^2 + B_h E_h^2 - 2D_h E_h F_h}{B_h C_h - F_h^2} - \frac{E_h^2}{C_h} \right) > 0. \]

(iii) By (6) and (30),
\[ \text{MSE}\left(\hat{C}_{yA_S}\right)_{\min} < \text{MSE}\left(\hat{C}_{yA_S^{(4)}}\right)_{\min} \text{ if } \]
\[ \sum_{h=1}^{L} W_h^2 C_{y_h}^2 \theta_h \left( B_h \left( E_h - C_h T_h \right) \right) \right) + F_h^2 \left( F_h T_h - D_h \right)^2 \right) > 0. \]

(iv) By (6) and (36),
\[ \text{MSE}\left(\hat{C}_{yA_S}\right)_{\min} < \text{MSE}\left(\hat{C}_{yA_S^{(5)}}\right)_{\min} \text{ if } \sum_{h=1}^{L} W_h^2 C_{y_h}^2 \theta_h \left( \frac{U_{1h} + S_{1h}}{Q_h} \right) > 0 \]
\[ U_{1h} = F_h^2 \left( E_h - F_h P_h \right)^2 + B_h C_h \left( D_h^2 - E_h^2 - 2F_h^2 P_h^2 \right) - 2C_h D_h P_h \left( B_h C_h - F_h^2 \right) \]
\[ S_{1h} = 2B_h E_h F_h \left( C_h P_h - D_h \right) + B_h^2 \left( E_h^2 + C_h^2 P_h^2 \right) \]
and
\[ Q_h = C_h \left( B_h C_h - F_h^2 \right). \]

(v) By (6) and (42),
\[ \text{MSE}\left(\hat{C}_{yA}\right)_{\min} < \text{MSE}\left(\hat{C}_{yA_{(6)}}\right) \text{ if } \]
\[ \sum_{h=1}^{L} W_h^2 C_{y_h}^2 \theta_h \left( \frac{U_{2h} + S_{2h}}{V_h} \right) > 0 \]
\[ U_{2h} = C_h \left( D_h - P_h B_h \right)^2 + B_h \left( E_h - T_h C_h \right)^2 \]
\[ S_{2h} = F_h^2 \left( 2 \left( P_h D_h + T_h E_h - P_h T_h F_h \right) - \left( P_h^2 B_h + T_h^2 C_h \right) \right) + 2F_h \left( P_h T_h B_h C_h - D_h E_h \right) \]
and
\[ V_h = B_h C_h - F_h^2. \]

(vi) By (6) and (48),
\[ \text{MSE}\left(\hat{C}_{yA_S}\right)_{\min} < \text{MSE}\left(\hat{C}_{yR_1S}\right) \text{ if } \]
\[
\sum_{h=1}^{L} W_h^2 C_j^2 \theta_h \left( \frac{C_h D_h^2 + B_h E_h^2 - 2D_h E_h F_h}{B_h C_h - F_h^2} - (2E_h - C_h) \right) > 0.
\]

(vii) By (6) and (54),

\[
\text{MSE}\left( \hat{\gamma}_{yA} \right)_{\text{min}} < \text{MSE}\left( \hat{\gamma}_{yR2} \right) \text{ if }
\]

\[
\sum_{h=1}^{L} W_h^2 C_j^2 \theta_h \left( \frac{C_h D_h^2 + B_h E_h^2 - 2D_h E_h F_h}{B_h C_h - F_h^2} - (2D_h - B_h) \right) > 0.
\]

(viii) By (6) and (60),

\[
\text{MSE}\left( \hat{\gamma}_{yA} \right)_{\text{min}} < \text{MSE}\left( \hat{\gamma}_{yRAS}^{(1)} \right) \text{ if }
\]

\[
\sum_{h=1}^{L} W_h^2 C_j^2 \theta_h \left( \frac{C_h D_h^2 + B_h E_h^2 - 2D_h E_h F_h}{B_h C_h - F_h^2} - M_h \right) > 0.
\]

(ix) By (6) and (66),

\[
\text{MSE}\left( \hat{\gamma}_{yA} \right)_{\text{min}} < \text{MSE}\left( \hat{\gamma}_{yRAS}^{(2)} \right) \text{ if }
\]

\[
\sum_{h=1}^{L} W_h^2 C_j^2 \theta_h \left( \frac{C_h D_h^2 + B_h E_h^2 - 2D_h E_h F_h}{B_h C_h - F_h^2} - O_h \right) > 0.
\]

The proposed estimator \( \hat{\gamma}_{yA} \) is more efficient than other existing estimators, if above conditions are satisfied.

**6. NUMERICAL ILLUSTRATION**

We use the following expression for efficiency comparison.

\[
\text{PRE}\left( \hat{\gamma}_r \right) = \frac{\text{MSE}\left( \hat{\gamma}_y \right)}{\text{MSE}\left( \hat{\gamma}_r \right)} \times 100,
\]

where \( r = A_s, W_{1s}, W_{2s}, R_{1s}, R_{2s}, RA_{s}^{(1)}, RA_{s}^{(2)} \)

**Data Set:** (Source: Murthy, 1967)

Let \( y = \) Output for 80 factories in a region
\( x = \) Number of workers and
\( z = \) Fixed capital

The data statistics are
\[ N = 80, \quad N_1 = 19, \quad N_2 = 32, \quad N_3 = 29, \quad n = 22, \quad n_1 = 11, \quad n_2 = 5, \quad n_3 = 6, \]
\[ W_1 = 0.24, \quad W_2 = 0.40, \quad W_3 = 0.3625, \quad \overline{Y}_1 = 2967.95, \quad \overline{Y}_2 = 4657.63, \quad \overline{Y}_3 = 7212.97, \]
\[ \bar{X}_1 = 65.16, \quad \bar{X}_2 = 139.97, \quad \bar{X}_3 = 589.41, \quad \bar{Z}_1 = 349.68, \quad \bar{Z}_2 = 706.59, \]
\[ \bar{Z}_3 = 2098.69, \quad S_{x_1} = 757.09, \quad S_{x_2} = 669.13, \quad S_{x_3} = 854.56, \quad S_{x_1} = 11.18, \]
\[ S_{x_2} = 44.36, \quad S_{x_3} = 226.42, \quad S_{c_2} = 109.45, \quad S_{c_2} = 109.22, \quad S_{c_3} = 637.71, \]
\[ \rho_{xy_1} = 0.81, \quad \rho_{xy_3} = 0.94, \quad \rho_{x_2} = 0.90, \quad \rho_{xy_2} = 0.89, \quad \rho_{yz_2} = 0.93, \quad \rho_{xz_2} = 0.85, \]
\[ \rho_{xy_2} = 0.98, \quad \rho_{yz_3} = 0.98, \quad \rho_{xz_3} = 0.97, \quad \lambda_{400_1} = 3.31, \quad \lambda_{400_3} = 1.53, \quad \lambda_{400_3} = 2.69, \]
\[ \lambda_{300_1} = -1.30, \quad \lambda_{300_2} = -0.04, \quad \lambda_{300_3} = 0.45, \quad \lambda_{220_1} = 1.41, \quad \lambda_{220_2} = 1.74, \]
\[ \lambda_{220_3} = 2.39, \quad \lambda_{210_1} = -0.68, \quad \lambda_{210_2} = 0.11, \quad \lambda_{210_3} = 0.50, \quad \lambda_{120_1} = -0.23, \]
\[ \lambda_{120_2} = 0.43, \quad \lambda_{120_3} = 0.53, \quad \lambda_{040_1} = 1.61, \quad \lambda_{040_2} = 3.08, \quad \lambda_{040_3} = 2.22, \]
\[ \lambda_{030_1} = 0.10, \quad \lambda_{030_2} = 0.88, \quad \lambda_{030_3} = 0.55, \quad \lambda_{004_1} = 2.40, \quad \lambda_{004_2} = 2.32, \]
\[ \lambda_{004_3} = 2.08, \quad \lambda_{000_1} = -0.79, \quad \lambda_{003_2} = 0.10, \quad \lambda_{003_3} = 0.18, \quad \lambda_{203_1} = -2.53, \]
\[ \lambda_{203_2} = 1.58, \quad \lambda_{202_3} = 2.32, \quad \lambda_{201_1} = -1.06, \quad \lambda_{201_2} = -0.04, \quad \lambda_{201_3} = 0.34, \]
\[ \lambda_{212_1} = -0.89, \quad \lambda_{102_2} = 0.002, \quad \lambda_{102_3} = 0.25, \quad \lambda_{022_3} = 1.45, \quad \lambda_{022_2} = 1.91, \]
\[ \lambda_{022_1} = 2.07, \quad \lambda_{012_1} = -0.39, \quad \lambda_{012_2} = 0.21, \quad \lambda_{012_3} = 0.30, \quad \lambda_{021_1} = -0.091, \]
\[ \lambda_{021_2} = 0.48, \quad \lambda_{021_3} = 0.42. \]

**Note:** It is assumed that the available budget \( C' = 3500 \) units and \( C_0' = 500 \) units. So that \( (C' - C_0') = 3000 \) units and we assumed that the costs of measurement \( c_h' \) in three strata are \( c_1' = 150, \quad c_2' = 170 \) and \( c_3' = 120 \) units respectively.

**Table 1:**

<table>
<thead>
<tr>
<th>Estimator</th>
<th>MSE</th>
<th>PRE</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \hat{C}_{yS}^{(1)} )</td>
<td>0.000327</td>
<td>100.000</td>
</tr>
<tr>
<td>( \hat{C}_{yR1S} )</td>
<td>0.000196</td>
<td>166.837</td>
</tr>
<tr>
<td>( \hat{C}_{yR2S} )</td>
<td>0.000130</td>
<td>251.538</td>
</tr>
<tr>
<td>( \hat{C}_{yW1S} )</td>
<td>0.000089</td>
<td>367.416</td>
</tr>
<tr>
<td>( \hat{C}_{yW2S} )</td>
<td>0.000150</td>
<td>218.000</td>
</tr>
<tr>
<td>( \hat{C}_{yRA1S}^{(1)} )</td>
<td>0.000450</td>
<td>72.667</td>
</tr>
<tr>
<td>( \hat{C}_{yRA1S}^{(2)} )</td>
<td>0.001105</td>
<td>29.593</td>
</tr>
<tr>
<td>( \hat{C}_{yAS} )</td>
<td>0.000077</td>
<td>424.675</td>
</tr>
</tbody>
</table>
In Table 1 we observed that a proposed estimators $\hat{\gamma}_{rA}$ is more efficient as compared to the other existing estimators.

7. CONCLUSION

We proposed an estimator for population coefficient of variation ($CV$) using two auxiliary variables. It is observed that a proposed estimator is more efficient than the usual sample estimator, ratio estimator, Walsh estimators and Archana and Rao estimators for $CV$.

REFERENCES

IMPACT OF EDUCATION ON GENDER DISCRIMINATION: A COMPARATIVE STUDY OF PUBLIC & PRIVATE ORGANIZATIONS OF PAKISTAN

Zahid Ali Channar\textsuperscript{1} Muneer-ud-Din Soomro\textsuperscript{2} and Arshad Haroon\textsuperscript{1}
\textsuperscript{1} Department of Commerce, Economics & Management Sciences, Isra University, Hyderabad, Pakistan. Email: zachanar@yahoo.com
\textsuperscript{2} Institute of Commerce, University of Sindh, Jamshoro, Pakistan. Email: muneeruddinsoomro@hotmail.com

ABSTRACT
This paper examines Gender Discrimination in workforce and the impact of demographic variable – education – on the gender discrimination. Close ended questionnaire was administered from 526 males and females of lower, middle and higher category employees of public and private health and education departments of Hyderabad and Jamshoro districts. Likert scale was used to measure the tendencies of respondents regarding gender discrimination. SPSS was used to conduct the Reliability tests of the continuous scale i.e. gender discrimination. Two-way ANOVA was used to assess the gender discrimination in different education groups of both organizations. The results of the test for public organizations were: the interaction effect between sex and education (sex*education) was F (5, 299) = 1.833, P = .106; main effect for sex was F (1,299) =27.064, P=0.000 and main effect for education was F (5,299) = 7.976, P=0.000. The results of the test for private organizations were: the interaction effect between sex and education (sex*education) was F (4, 183) =1.440, P = .222; main effect for sex was F (1,183) = 4.028, P=0.046 and main effect for education was F (5,183) = 11.444, P=0.000. The result of interaction effect meant that the influence of education on gender discrimination was not different for males and females in both organizations. Result of main effect for sex meant that females were discriminated more than males again in both organizations. Two-way ANOVA for education showed that the females having lower qualification were discriminated more than the females having higher qualification, in both public and private organizations.

KEYWORDS:
Equal Employment Opportunity; Gender Discrimination; Workforce; Demographic variable.

1. INTRODUCTION
Equal Employment Opportunity (EEO) is a condition in which all individuals have an equal chance of employment regardless of their race, color, sex and religion. EEO prohibits all forms of employment discrimination. Affirmative Action (AA) is an action taken to overcome the effects of past or present practices, policies or other barriers to EEO. If EEO is not acted upon properly, it will result in “Gender Discrimination”.

609
Workforce was strikingly homogenous i.e. workers were alike (belonging to one race, sex, having same qualification etc) upto 1950s. Therefore they shared the same interests and needs. Personnel’s job was, therefore, simpler. In the United States, Passage of federal legislation in the 1960s prohibited employment discrimination. This resulted in the heterogeneous workforce that had varied personal characteristics and thus evolved the term’ workforce diversity’.

Wayne (1995) has explained that Managing Diversity means establishing a heterogeneous work force to perform to its potentials in an equitable work environment, where no member or group of members has an advantage or a disadvantage. Managing diversity is very essential for any organization, especially in this era of globalization. Managing diversity is not only a hot topic in HRM but also in Organizational Behavior (OB). Wayne (1995) writes, as a society, we espouse equality of opportunity, rather than equality of outcomes. That is, the broad goal is to provide for all citizens of the country, regardless of race, age, gender, religion, national origin, or disability, an equal opportunity to compete for the jobs for which they are qualified. The objective, therefore, is EEO, not EE (Equal Employment, or equal number of employees from various sub groups). With this we can achieve the nation’s goals of equality of opportunity, full participation, independent living, and economic reliance. John (2004) says that throughout history, western society has accepted the principle that people should be rewarded according to the worth of their contribution.

2. STATEMENT OF THE PROBLEM

Not only the lowly educated females but also highly educated females confront discrimination, which hinder their career not only through sticky floor but also through glass ceiling effects.

3. OBJECTIVE OF THE STUDY

To study discrimination against females at workforce, at different qualification levels.

4. HYPOTHESIS

The effect of education on gender discrimination in workforce is not more for females than males in public organizations.

5. SCOPE OF THE STUDY

This research assessed Gender Discrimination in different education groups of both sexes in both public and private organizations of Hyderabad and Jamshoro districts. Health and Education departments were taken from each of the above sectors.

6. GENDER DISCRIMINATION IN DIFFERENT LEVELS OF EDUCATION IN PUBLIC ORGANIZATIONS

Here the impact of Education and Sex on the gender discrimination in work force is assessed. Variables used here are:

- One continuous independent variable (Gender Discrimination in workforce).
Two categorical variables (Sex: Male/Female; Education: Group 1: Matric = 1; Group 2: Intermediate = 2; Group 3: Graduate / Diploma = 3; Group 4: Masters = 4; Group 5: M.Phil = 5; Group 6: Ph.D = 6).

Table I: Descriptive statistics for two-way ANOVA for education

<table>
<thead>
<tr>
<th>Sex</th>
<th>Education</th>
<th>Mean</th>
<th>Std. Deviation</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>matric</td>
<td>17.5000</td>
<td>3.39116</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>intermediate</td>
<td>22.3333</td>
<td>2.50333</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>graduate / diploma</td>
<td>20.8958</td>
<td>3.07522</td>
<td>48</td>
</tr>
<tr>
<td></td>
<td>masters</td>
<td>19.1622</td>
<td>5.20544</td>
<td>74</td>
</tr>
<tr>
<td></td>
<td>M.Phil</td>
<td>18.8889</td>
<td>4.67559</td>
<td>9</td>
</tr>
<tr>
<td></td>
<td>Ph.D</td>
<td>17.7500</td>
<td>3.37004</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>19.6821</td>
<td>4.44428</td>
<td>151</td>
</tr>
<tr>
<td>Female</td>
<td>matric</td>
<td>24.8667</td>
<td>3.70071</td>
<td>15</td>
</tr>
<tr>
<td></td>
<td>intermediate</td>
<td>27.1667</td>
<td>4.87511</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>graduate / diploma</td>
<td>25.2857</td>
<td>3.32290</td>
<td>49</td>
</tr>
<tr>
<td></td>
<td>masters</td>
<td>21.1733</td>
<td>4.82777</td>
<td>75</td>
</tr>
<tr>
<td></td>
<td>M.Phil</td>
<td>21.8000</td>
<td>4.32435</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>Ph.D</td>
<td>20.5000</td>
<td>4.06202</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>22.9812</td>
<td>4.70577</td>
<td>160</td>
</tr>
<tr>
<td>Total</td>
<td>matric</td>
<td>22.7619</td>
<td>4.90820</td>
<td>21</td>
</tr>
<tr>
<td></td>
<td>intermediate</td>
<td>24.7500</td>
<td>4.47468</td>
<td>12</td>
</tr>
<tr>
<td></td>
<td>graduate / diploma</td>
<td>23.1134</td>
<td>3.87534</td>
<td>97</td>
</tr>
<tr>
<td></td>
<td>masters</td>
<td>20.1745</td>
<td>5.10264</td>
<td>149</td>
</tr>
<tr>
<td></td>
<td>M.Phil</td>
<td>19.9286</td>
<td>4.61543</td>
<td>14</td>
</tr>
<tr>
<td></td>
<td>Ph.D</td>
<td>19.2778</td>
<td>3.92287</td>
<td>18</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>21.3794</td>
<td>4.86237</td>
<td>311</td>
</tr>
</tbody>
</table>

Table II: Levene’s Test for Education

<table>
<thead>
<tr>
<th>Levene's Test of Equality of Error Variancesa</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dependent Variable: Discrimination in work force</td>
</tr>
<tr>
<td>F</td>
</tr>
<tr>
<td>2.238</td>
</tr>
</tbody>
</table>

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.

a. Design: Intercept + Sex + Education + Sex * Education
Table III: Tests of between subjects effects for education

Dependent Variable: Discrimination in work force

<table>
<thead>
<tr>
<th>Source</th>
<th>Type III Sum of Squares</th>
<th>df</th>
<th>Mean Square</th>
<th>F</th>
<th>Sig.</th>
<th>Partial Eta Squared</th>
</tr>
</thead>
<tbody>
<tr>
<td>Corrected Model</td>
<td>1774.860a</td>
<td>11</td>
<td>161.351</td>
<td>8.686</td>
<td>.000</td>
<td>.242</td>
</tr>
<tr>
<td>Intercept</td>
<td>56552.084</td>
<td>1</td>
<td>56552.084</td>
<td>3.044E3</td>
<td>.000</td>
<td>.911</td>
</tr>
<tr>
<td>Sex</td>
<td>502.750</td>
<td>1</td>
<td>502.750</td>
<td>27.064</td>
<td>.000</td>
<td>.083</td>
</tr>
<tr>
<td>Education</td>
<td>740.804</td>
<td>5</td>
<td>148.161</td>
<td>7.976</td>
<td>.000</td>
<td>.118</td>
</tr>
<tr>
<td>Sex * Education</td>
<td>170.297</td>
<td>5</td>
<td>34.059</td>
<td>1.833</td>
<td>.106</td>
<td>.030</td>
</tr>
<tr>
<td>Error</td>
<td>5554.369</td>
<td>299</td>
<td>18.576</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>149481.000</td>
<td>311</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Corrected Total</td>
<td>7329.228</td>
<td>310</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

a. R Squared = .242 (Adjusted R Squared = .214)

Table IV: Multiple comparisons for education

Discrimination in work force

Tukey HSD

<table>
<thead>
<tr>
<th>(I) Education</th>
<th>(J) Education</th>
<th>Mean Difference (I-J)</th>
<th>Std. Error</th>
<th>Sig.</th>
<th>95% Confidence Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>matric</td>
<td>Intermediate</td>
<td>-1.9881</td>
<td>1.55969</td>
<td>.799</td>
<td>-6.4618 - 2.4856</td>
</tr>
<tr>
<td></td>
<td>graduate/diploma</td>
<td>-.3515</td>
<td>1.03735</td>
<td>.999</td>
<td>-3.3269 - 2.6239</td>
</tr>
<tr>
<td></td>
<td>Masters</td>
<td>2.5874</td>
<td>1.00462</td>
<td>.107</td>
<td>-2.942 - 5.4690</td>
</tr>
<tr>
<td></td>
<td>MPhil</td>
<td>2.8333</td>
<td>1.48711</td>
<td>.401</td>
<td>-1.833 - 7.0988</td>
</tr>
<tr>
<td></td>
<td>PhD</td>
<td>3.4841</td>
<td>1.38442</td>
<td>.123</td>
<td>-4.868 - 7.4551</td>
</tr>
<tr>
<td>intermediate</td>
<td>matric</td>
<td>1.9881</td>
<td>1.55969</td>
<td>.799</td>
<td>-2.4856 - 6.4618</td>
</tr>
<tr>
<td></td>
<td>graduate/diploma</td>
<td>1.6366</td>
<td>1.31892</td>
<td>.816</td>
<td>-2.1465 - 5.4197</td>
</tr>
<tr>
<td></td>
<td>masters</td>
<td>4.5755*</td>
<td>1.29333</td>
<td>.006</td>
<td>.8658 - 8.2852</td>
</tr>
<tr>
<td></td>
<td>MPhil</td>
<td>4.8214</td>
<td>1.69556</td>
<td>.054</td>
<td>-.0419 - 9.6848</td>
</tr>
<tr>
<td></td>
<td>PhD</td>
<td>5.4722*</td>
<td>1.60626</td>
<td>.010</td>
<td>.8650 - 10.0795</td>
</tr>
<tr>
<td>graduate/</td>
<td>matric</td>
<td>.3515</td>
<td>1.03735</td>
<td>.999</td>
<td>-2.6239 - 3.3269</td>
</tr>
<tr>
<td>diploma</td>
<td>intermediate</td>
<td>-1.6366</td>
<td>1.31892</td>
<td>.816</td>
<td>-5.4197 - 2.1465</td>
</tr>
<tr>
<td></td>
<td>masters</td>
<td>2.9389*</td>
<td>.56230</td>
<td>.000</td>
<td>1.3261 - 4.5518</td>
</tr>
<tr>
<td></td>
<td>MPhil</td>
<td>3.1848</td>
<td>1.23223</td>
<td>.104</td>
<td>-3.3496 - 7.6192</td>
</tr>
<tr>
<td></td>
<td>PhD</td>
<td>3.8356*</td>
<td>1.10614</td>
<td>.008</td>
<td>.6629 - 7.0084</td>
</tr>
<tr>
<td>masters</td>
<td>matric</td>
<td>-2.5874</td>
<td>1.00462</td>
<td>.107</td>
<td>-5.4690 - 2.942</td>
</tr>
<tr>
<td></td>
<td>intermediate</td>
<td>-4.5755*</td>
<td>1.29333</td>
<td>.006</td>
<td>-8.2852 - 3.8658</td>
</tr>
<tr>
<td></td>
<td>graduate/diploma</td>
<td>-2.9389*</td>
<td>.56230</td>
<td>.000</td>
<td>-4.5518 - 1.3261</td>
</tr>
<tr>
<td></td>
<td>MPhil</td>
<td>.2459</td>
<td>1.20481</td>
<td>1.000</td>
<td>-3.2098 - 3.7017</td>
</tr>
<tr>
<td></td>
<td>PhD</td>
<td>.8967</td>
<td>1.07550</td>
<td>.961</td>
<td>-2.1881 - 3.9816</td>
</tr>
<tr>
<td>MPhil</td>
<td>matric</td>
<td>-2.8333</td>
<td>1.48711</td>
<td>.401</td>
<td>-7.0988 - 1.4321</td>
</tr>
<tr>
<td></td>
<td>intermediate</td>
<td>-4.8214</td>
<td>1.69556</td>
<td>.054</td>
<td>-9.6848 - .0419</td>
</tr>
<tr>
<td></td>
<td>graduate/diploma</td>
<td>-3.1848</td>
<td>1.23223</td>
<td>.104</td>
<td>-6.7192 - .3496</td>
</tr>
<tr>
<td></td>
<td>masters</td>
<td>-.2459</td>
<td>1.20481</td>
<td>1.000</td>
<td>-3.7017 - 3.2098</td>
</tr>
<tr>
<td></td>
<td>PhD</td>
<td>.6508</td>
<td>1.53588</td>
<td>.998</td>
<td>-3.7546 - 5.0561</td>
</tr>
<tr>
<td>PhD</td>
<td>matric</td>
<td>-3.4841</td>
<td>1.38442</td>
<td>.123</td>
<td>-7.4551 - .4868</td>
</tr>
<tr>
<td></td>
<td>intermediate</td>
<td>-5.4722*</td>
<td>1.60626</td>
<td>.010</td>
<td>-10.0795 - .8650</td>
</tr>
<tr>
<td></td>
<td>graduate/diploma</td>
<td>-3.8356*</td>
<td>1.10614</td>
<td>.008</td>
<td>-7.0084 - -6.629</td>
</tr>
<tr>
<td></td>
<td>masters</td>
<td>-.8967</td>
<td>1.07550</td>
<td>.961</td>
<td>-3.9816 - 2.1881</td>
</tr>
<tr>
<td></td>
<td>MPhil</td>
<td>-.6508</td>
<td>1.53588</td>
<td>.998</td>
<td>-5.0561 - 3.7546</td>
</tr>
</tbody>
</table>
A two-way between-groups analysis of variance was conducted to explore the impact of sex and education on gender discrimination in work force. Subjects were divided into six education groups (Group 1: Matriculation; Group 2: Intermediate; Group 3: Graduate / Diploma; Group 4: Masters; Group 5: M.Phil; Group 6: Ph.D). The interaction effect between sex and education was not statistically significant, $F(5, 299) = 1.833, p = .106$. There was a statistically significant main effect for education, $F(5, 299) = 7.976, p = .000$. Effect size for education was moderate ($\eta^2 = .118$). The main effect for sex, $F(1, 299) = 27.064, p = .000$ was also significant with moderate effect size ($\eta^2 = .083$).

In the above results ‘interaction effect’ for Sex*Education is more than .05, which means that the influence of education on gender discrimination is not different for males and females. ‘Main effect’ for sex is .000, which means that males and females differ in terms of their gender discrimination scores [females are discriminated more ($m=22.981$) than males ($m=19.682$)]. ‘Main effect’ for education is .000, which means that education groups, intermediate group ($M = 24.750, SD = 4.474$) was significantly different from the masters group ($M = 20.174, SD = 5.102$) and Ph.D group ($M = 19.277, SD = 3.922$). The mean score for the graduate / diploma group ($M = 23.113, SD = 3.875$) was significantly different from the masters group ($M = 20.174, SD = 5.102$) and Ph.D group ($M = 19.277, SD = 3.922$). Effect size for sex is .083, which is medium which means that the actual difference in the mean values for males (19.682) and females (22.981) is moderate. Effect size for education is .118, which is large.

![Profile Plots](image)

Fig. I: Graph of discrimination in different education groups of males & females
size for education is .118, which is also medium. It means the actual difference in the mean values of education groups is moderate i.e. 22.761, 24.750, 23.113, 20.174, 19.928, and 19.277.

7. GENDER DISCRIMINATION IN DIFFERENT LEVELS OF EDUCATION IN PRIVATE ORGANIZATIONS

Here the impact of Education and Sex on the gender discrimination in workforce is assessed. Variables used here are:
- One continuous independent variable (Gender Discrimination in workforce).
- Two categorical variables (Sex: Male/Female; Education: Group 1: Matriculation = 1; Group 2: Intermediate = 2; Group 3: Graduate / Diploma = 3; Group 4: Masters = 4; Group 5: M.Phil = 5; Group 6: Ph.D = 6).

Table V: Descriptive statistics for two-way ANOVA for education
Dependent Variable: Discrimination in workforce

<table>
<thead>
<tr>
<th>Sex</th>
<th>Education</th>
<th>Mean</th>
<th>Std. Deviation</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>matriculation</td>
<td>23.8889</td>
<td>2.57121</td>
<td>9</td>
</tr>
<tr>
<td></td>
<td>intermediate</td>
<td>24.3333</td>
<td>2.51661</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>graduate / diploma</td>
<td>20.7857</td>
<td>3.45722</td>
<td>28</td>
</tr>
<tr>
<td></td>
<td>masters</td>
<td>18.5588</td>
<td>3.22106</td>
<td>34</td>
</tr>
<tr>
<td></td>
<td>MPhil</td>
<td>16.8889</td>
<td>4.75511</td>
<td>9</td>
</tr>
<tr>
<td></td>
<td>PhD</td>
<td>16.0000</td>
<td>3.67423</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>19.6932</td>
<td>4.01823</td>
<td>88</td>
</tr>
<tr>
<td>Female</td>
<td>matriculation</td>
<td>26.7273</td>
<td>3.52394</td>
<td>11</td>
</tr>
<tr>
<td></td>
<td>intermediate</td>
<td>24.7000</td>
<td>3.65300</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td>graduate / diploma</td>
<td>25.7812</td>
<td>4.48463</td>
<td>32</td>
</tr>
<tr>
<td></td>
<td>masters</td>
<td>20.5769</td>
<td>5.04624</td>
<td>52</td>
</tr>
<tr>
<td></td>
<td>MPhil</td>
<td>18.0000</td>
<td>.</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>23.1509</td>
<td>5.27671</td>
<td>106</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Sex</th>
<th>Education</th>
<th>Mean</th>
<th>Std. Deviation</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total</td>
<td>matriculation</td>
<td>25.4500</td>
<td>3.37912</td>
<td>20</td>
</tr>
<tr>
<td></td>
<td>intermediate</td>
<td>24.6154</td>
<td>3.33013</td>
<td>13</td>
</tr>
<tr>
<td></td>
<td>graduate / diploma</td>
<td>23.4500</td>
<td>4.72794</td>
<td>60</td>
</tr>
<tr>
<td></td>
<td>masters</td>
<td>19.7791</td>
<td>4.50464</td>
<td>86</td>
</tr>
<tr>
<td></td>
<td>MPhil</td>
<td>17.0000</td>
<td>4.49691</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td>PhD</td>
<td>16.0000</td>
<td>3.67423</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>21.5825</td>
<td>5.04035</td>
<td>194</td>
</tr>
</tbody>
</table>

Table VI: Levene’s test for education

Levene’s Test of Equality of Error Variances

<p>| | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>F</td>
<td>df1</td>
<td>df2</td>
<td>Sig.</td>
<td></td>
</tr>
<tr>
<td>1.399</td>
<td>10</td>
<td>183</td>
<td>.184</td>
<td></td>
</tr>
</tbody>
</table>

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.

a. Design: Intercept + Sex + Education + Sex * Education
Table VII: Tests of between subjects effects for education

<table>
<thead>
<tr>
<th>Source</th>
<th>Type III Sum of Squares</th>
<th>df</th>
<th>Mean Square</th>
<th>F</th>
<th>Sig.</th>
<th>Partial Eta Squared</th>
</tr>
</thead>
<tbody>
<tr>
<td>Corrected Model</td>
<td>1771.196*</td>
<td>10</td>
<td>177.120</td>
<td>10.349</td>
<td>.000</td>
<td>.361</td>
</tr>
<tr>
<td>Intercept</td>
<td>24402.536</td>
<td>1</td>
<td>24402.536</td>
<td>1.426E3</td>
<td>.000</td>
<td>.886</td>
</tr>
<tr>
<td>Sex</td>
<td>68.936</td>
<td>1</td>
<td>68.936</td>
<td>4.028</td>
<td>.046</td>
<td>.022</td>
</tr>
<tr>
<td>Education</td>
<td>979.288</td>
<td>5</td>
<td>195.858</td>
<td>11.444</td>
<td>.000</td>
<td>.238</td>
</tr>
<tr>
<td>Sex * Education</td>
<td>98.606</td>
<td>4</td>
<td>24.651</td>
<td>1.440</td>
<td>.222</td>
<td>.031</td>
</tr>
<tr>
<td>Error</td>
<td>3131.984</td>
<td>183</td>
<td>17.115</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>95269.000</td>
<td>194</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Corrected Total</td>
<td>4903.180</td>
<td>193</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* R Squared = .361 (Adjusted R Squared = .326)

Table VIII: Multiple comparisons for education

Based on observed means.

<table>
<thead>
<tr>
<th>(I) Education</th>
<th>(J) Education</th>
<th>Mean Difference (I-J)</th>
<th>Std. Error</th>
<th>Sig.</th>
<th>95% Confidence Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>matric</td>
<td>graduate / diploma</td>
<td>.8346</td>
<td>1.47385</td>
<td>.993</td>
<td>-3.4103</td>
</tr>
<tr>
<td>matric</td>
<td>masters</td>
<td>2.0000</td>
<td>1.06817</td>
<td>.423</td>
<td>-1.0765</td>
</tr>
<tr>
<td>matric</td>
<td>MPhil</td>
<td>5.6709*</td>
<td>1.02701</td>
<td>.000</td>
<td>2.7130</td>
</tr>
<tr>
<td>matric</td>
<td>PhD</td>
<td>8.4500*</td>
<td>1.60225</td>
<td>.000</td>
<td>3.8353</td>
</tr>
<tr>
<td>matric</td>
<td>PhD</td>
<td>9.4500*</td>
<td>2.06849</td>
<td>.000</td>
<td>3.4924</td>
</tr>
<tr>
<td>matric</td>
<td>masters</td>
<td>1.1654</td>
<td>1.26561</td>
<td>.941</td>
<td>-2.4798</td>
</tr>
<tr>
<td>matric</td>
<td>MPhil</td>
<td>4.8363*</td>
<td>1.23106</td>
<td>.002</td>
<td>1.2907</td>
</tr>
<tr>
<td>matric</td>
<td>PhD</td>
<td>7.6154*</td>
<td>1.74011</td>
<td>.000</td>
<td>2.6036</td>
</tr>
<tr>
<td>matric</td>
<td>PhD</td>
<td>8.6154*</td>
<td>2.17703</td>
<td>.001</td>
<td>3.4924</td>
</tr>
<tr>
<td>intermediate</td>
<td>graduate / diploma</td>
<td>-.8346</td>
<td>1.47385</td>
<td>.993</td>
<td>-5.0795</td>
</tr>
<tr>
<td>intermediate</td>
<td>masters</td>
<td>1.1654</td>
<td>1.26561</td>
<td>.941</td>
<td>-2.4798</td>
</tr>
<tr>
<td>intermediate</td>
<td>MPhil</td>
<td>4.8363*</td>
<td>1.23106</td>
<td>.002</td>
<td>1.2907</td>
</tr>
<tr>
<td>intermediate</td>
<td>PhD</td>
<td>7.6154*</td>
<td>1.74011</td>
<td>.000</td>
<td>2.6036</td>
</tr>
<tr>
<td>intermediate</td>
<td>PhD</td>
<td>8.6154*</td>
<td>2.17703</td>
<td>.001</td>
<td>3.4924</td>
</tr>
<tr>
<td>graduate / diploma</td>
<td>matric</td>
<td>-2.0000</td>
<td>1.06817</td>
<td>.423</td>
<td>-5.0765</td>
</tr>
<tr>
<td>graduate / diploma</td>
<td>intermediate</td>
<td>-1.1654</td>
<td>1.26561</td>
<td>.941</td>
<td>-4.8105</td>
</tr>
<tr>
<td>graduate / diploma</td>
<td>masters</td>
<td>3.6709*</td>
<td>.69588</td>
<td>.002</td>
<td>1.6667</td>
</tr>
<tr>
<td>graduate / diploma</td>
<td>MPhil</td>
<td>6.4500*</td>
<td>1.41305</td>
<td>.000</td>
<td>2.3802</td>
</tr>
<tr>
<td>graduate / diploma</td>
<td>PhD</td>
<td>7.4500*</td>
<td>1.92566</td>
<td>.002</td>
<td>1.9038</td>
</tr>
<tr>
<td>masters</td>
<td>matric</td>
<td>-5.6709*</td>
<td>1.02701</td>
<td>.000</td>
<td>-8.6289</td>
</tr>
<tr>
<td>masters</td>
<td>intermediate</td>
<td>-4.8363*</td>
<td>1.23106</td>
<td>.002</td>
<td>-8.3820</td>
</tr>
<tr>
<td>masters</td>
<td>graduate / diploma</td>
<td>-3.6709*</td>
<td>.69588</td>
<td>.002</td>
<td>-5.6752</td>
</tr>
<tr>
<td>masters</td>
<td>MPhil</td>
<td>2.7791</td>
<td>1.38220</td>
<td>.340</td>
<td>-1.2019</td>
</tr>
<tr>
<td>masters</td>
<td>PhD</td>
<td>7.4500*</td>
<td>1.92566</td>
<td>.002</td>
<td>1.9038</td>
</tr>
<tr>
<td>masters</td>
<td>PhD</td>
<td>-2.7791</td>
<td>1.38220</td>
<td>.340</td>
<td>-6.7600</td>
</tr>
<tr>
<td>MPhil</td>
<td>matric</td>
<td>-8.4500*</td>
<td>1.60225</td>
<td>.000</td>
<td>-13.0647</td>
</tr>
<tr>
<td>MPhil</td>
<td>intermediate</td>
<td>-7.6154*</td>
<td>1.74011</td>
<td>.000</td>
<td>-12.6272</td>
</tr>
<tr>
<td>MPhil</td>
<td>graduate / diploma</td>
<td>-6.4500*</td>
<td>1.41305</td>
<td>.000</td>
<td>-10.5198</td>
</tr>
<tr>
<td>MPhil</td>
<td>masters</td>
<td>-2.7791</td>
<td>1.38220</td>
<td>.340</td>
<td>-6.7600</td>
</tr>
<tr>
<td>MPhil</td>
<td>PhD</td>
<td>1.0000</td>
<td>2.26592</td>
<td>.998</td>
<td>-5.5262</td>
</tr>
<tr>
<td>PhD</td>
<td>matric</td>
<td>-9.4500*</td>
<td>2.06849</td>
<td>.000</td>
<td>-15.4076</td>
</tr>
<tr>
<td>PhD</td>
<td>intermediate</td>
<td>-8.6154*</td>
<td>2.17703</td>
<td>.001</td>
<td>-14.8856</td>
</tr>
<tr>
<td>PhD</td>
<td>graduate / diploma</td>
<td>-7.4500*</td>
<td>1.92566</td>
<td>.002</td>
<td>-12.9926</td>
</tr>
<tr>
<td>PhD</td>
<td>masters</td>
<td>-3.7791</td>
<td>1.90314</td>
<td>.355</td>
<td>-9.2604</td>
</tr>
<tr>
<td>PhD</td>
<td>MPhil</td>
<td>-1.0000</td>
<td>2.26592</td>
<td>.998</td>
<td>-7.5262</td>
</tr>
</tbody>
</table>

The error term is Mean Square (Error) = 17.115.

* The mean difference is significant at the .05 level.
A two-way between-groups analysis of variance was conducted to explore the impact of sex and education on gender discrimination in work force. Subjects were divided into six education groups (Group 1: Matric; Group 2: Intermediate; Group 3: Graduate / Diploma; Group 4: Masters; Group 5: M.Phil; Group 6: Ph.D). The interaction effect between sex and education was not statistically significant, $F(4, 183) = 1.440, p = .222$. There was a statistically significant main effect for education, $F(5, 183) = 11.444, p = .000$. Effect size for education was large (partial eta squared = .238). The main effect for sex, $F(1, 183) = 4.028, p = .046$ was also significant with small effect size (partial eta squared = .022).

In the above results ‘interaction effect’ for Sex*Education is more than .05, which means that the influence of education on gender discrimination is not different for males and females. ‘Main effect’ for sex is .046, which means that males and females differ in terms of their gender discrimination scores [females are discriminated more ($m=23.150$) than males ($m=19.693$)]. ‘Main effect’ for education is .000, which means that education groups, matric group ($M = 25.450, SD = 3.379$) was significantly different from the masters group ($M = 19.779, SD = 4.504$), M.Phil group ($M = 17.000, SD = 4.496$) and PhD group ($M = 16.000, SD = 3.674$). The mean score for the intermediate group ($M = 24.615, SD = 3.330$) was also significantly different from the masters group ($M = 19.779, SD = 4.504$), M.Phil group ($M = 17.000, SD = 4.496$) and Ph.D group ($M = 16.000, SD = 3.674$). The mean score for the graduate / diploma group ($M = 23.450, SD = 4.727$) was also significantly different from the masters group ($M = 19.779, SD = 4.504$), M.Phil
group (M = 17.000, SD = 4.496) and Ph.D group (M = 16.000, SD = 3.674). The mean score for the masters group (M = 19.779, SD = 4.504) did not significantly differ significantly from M.Phil group (M = 17.000, SD = 4.496) and Ph.D group (M = 16.000, SD = 3.674). Effect size for sex is .022, which is small. It means though the effect has reached statistical significance; the actual difference in the mean values for males (m=19.693) and females (23.150) is small. Effect size for education is .238, which is large. It means that the actual difference in the mean values of the education groups is large i.e., 25.450, 24.615, 23.450, 19.779, 17.000 and 16.000.

8. COMMENTS & CONCLUSIONS

Hypothesis 15 inferred that “the effect of education on gender discrimination in workforce is different for males and females in public organizations than private organizations”. This hypothesis was also tested by two-way ANOVA. The results of the test for public organizations were: the interaction effect between sex and education (sex*education) was F (5, 299) = 1.833, P = .106; main effect for sex was F (1,299) =27.064, P=0.000 and main effect for education was F (5,299) = 7.976, P=0.000. The result of interaction effect meant that the influence of education on gender discrimination was not different for males and females. Result of main effect for sex meant that females were discriminated more (mean = 22.981) than males (mean = 19.682). Result of main effect for education indicated that females were discriminated more than males in majority of the education groups. As the result for the interaction effect between sex and education was statistically not significant, therefore we accept the null hypothesis and reject the alternative hypothesis:

\( H_0: \) The effect of education on gender discrimination in workforce is not more for females than males in public organizations.

\( H_A: \) The effect of education on gender discrimination in workforce is more for females than males in public organizations.

The results of the test for private organizations were: the interaction effect between sex and education (sex*education) was F (4, 183) =1.440, P = .222; main effect for sex was F (1,183) = 4.028, P=0.046 and main effect for education was F (5,183) = 11.444, P=0.000. The result of interaction effect meant that the influence of education on gender discrimination was not different for males and females. Result of main effect for sex meant that females were discriminated more (mean = 23.150) than males (mean = 19.693). Result of main effect for education indicated that females were discriminated more than males in majority of the education groups. As the result for the interaction effect between sex and education was statistically not significant, therefore we accept the null hypothesis and reject the alternative hypothesis:

\( H_0: \) The effect of education on gender discrimination in workforce is not more for females than males in public organizations.

\( H_A: \) The effect of education on gender discrimination in workforce is more for females than males in public organizations.

We concluded from above results that the effect of education on gender discrimination in workforce was not different for females and males in public organizations than private organizations.
REFERENCES

STATISTICAL STUDY OF AGRICULTURE DEVELOPMENT IN PAKISTAN

Ambreen Zeb Khaskhelly¹ and Imran Anwar Ujan²

¹ Department of Economics, University of Sindh, Jamshoro, Pakistan. Email: ambrzk_khas@hotmail.com
² Institute of Information & Communication Technology, University of Sindh, Jamshoro, Pakistan. Email: iujan@yahoo.com

ABSTRACT

Agriculture is the single largest sector of Pakistan’s economy. It contributes 20.9% of the GDP, employs 70% of the workforce and is the major source of foreign exchange earnings. The major objectives of this paper are to document and examine the performance of agriculture in various sectors of the economy, identify the problems that affect growth and performance and recommend strategies for sustainability of agricultural sector in the next decade. The paper analyzed is 1950-55 to 2002-2003. The major conclusion drawn from this paper production shows that between 1950-55 and 2002-03 for all major as well as minor crops, there has been increase almost in total output over the years. The implementation of green revolution had created employment opportunities in crop production sub-sector. Empirical evidence on land reforms establishes that net incomes of tenants have increased faster than those of landlords. The impact of mechanisation on income distribution shows that only big farmers benefited from the tractor tube well technology. Pakistan faces serious problems of its forest reserve, if this situation continues; Pakistan will loose most of its forest in the next 40 years. The impact of commodities like tea, edible oil, milk production is a significant burden on foreign exchange reserves.

This paper recommends some measures as a positive step towards addressing the above problems so as to improve the performance of the agricultural sector.

INTRODUCTION

Agriculture is the single largest sector of Pakistan’s economy. It contributes 20.9% of the GDP, employs 70% of the workforce and is the major source of foreign exchange earnings. About 68% of the population lives in rural Pakistan and depends upon agriculture for subsistence (Malik, 2005).

The major crops such as wheat, cotton, rice, sugarcane and maize account for 41% of the value added and minor crops 10% in overall agriculture. It accounts for 37% of agriculture value added and about 9.4% of the GDP. Similarly, fisheries play an important role in national income through export earnings.

Government has identified agriculture as a priority area for addressing problems of unemployment, poverty alleviation and for fostering economic development. The agricultural policy focuses on sustainable food security increasing productivity,
commercial agriculture, income substitution, income diversification and exports orientation (Saheed, 2010).

The overall policy goal is to raise productivity and profitability of the farming community enabling the country to raise living standard of rural masses. The policy is directed to ensure that the natural resource base as land and watering is conserved and the situation of short water supply or draught is mitigated (Saeed, 2005).

In respect of farm inputs attempt has been made to arrange adulteration free availability at reasonable prices, steps have been taken to reduce cost of farm inputs. In the procurement of agricultural commodities, the role of private sector is being emphasised leading to public private sector partnership.

The institutional services as agriculture research and extension are geared to raise productivity and profitability of the farmer and address issues faced by the farming community at field level. The services in agricultural marketing are being renovated, it is planned to establish new markets in areas where marketing structure is weak (Khan, 2010).

Agriculture in Pakistan is closely linked to the rest of the country. It supplies a regular flow of workers to non-agricultural sector and is taxed to provide cheap food to the urban workers. Due to this trend, the four largest cities have more than doubled their population and domestic prices of food have been lower than international prices.

Agricultural crops such as cotton, sugarcane constitute raw materials for two of the most important industries in Pakistan namely textile and sugar. On the demand side, it consumes fertiliser manufactured domestically and is an increasingly important consumer of electricity and engineering goods (Government of Pakistan 2008 – 09).

PROBLEM STATEMENTS

The performance of agricultural sector in Pakistan has been unsatisfactory. It is chiefly attributable to traditional method of cultivation, lack of motivation on the part of our illiterate and uninformed rural population and more importantly to natural disasters like floods and pest attacks. Recent evidence reveals a dismal picture of agricultural sector in Pakistan. From input output analysis conducted by analysts, it was concluded that the agricultural sector has had a decreasing returns on these inputs. These decreasing returns cannot be termed as overall decreasing returns because the supplies of land and water are much less elastic and cannot be increased.

OBJECTIVES

The objectives of the paper are:
  i) To examine agricultural production, green revolution and land reforms.
  ii) To assess the impact of farm mechanisation on income, productivity and employment.
  iii) To analyse forestry and fisheries issues.
  iv) To examine livestock and poultry issues.
  v) To evaluate rural development programmes and their impact on rural communities.
Production of major Crops
Total production shows that between 1950 – 1955 and 2002 – 2003 for all major as well as minor crops, there has been a very substantial increase almost in total output over the. Although overall output for major crops has increased, what is discerning is that the rate of increase, the average annual growth rate of major crops since 1980 has fallen. In the decade 1980 – 1990, the average annual growth of major crops was 3.34%, which fell to around 2.4% in the 1990s. While lower rates of increases in output and area were disturbing factor, so too is the instability and variability in output from year to year.

Table 1:
Area under Major Crops 1950 – 2003 (‘000 Hectares)

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>Minimum</th>
<th>Maximum</th>
<th>Mean</th>
<th>Std. Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wheat</td>
<td>11</td>
<td>4154.0</td>
<td>8306.6</td>
<td>6.427E3</td>
<td>1472.7283</td>
</tr>
<tr>
<td>Rice</td>
<td>11</td>
<td>947.0</td>
<td>2333.8</td>
<td>1.692E3</td>
<td>480.3408</td>
</tr>
<tr>
<td>Sugarcane</td>
<td>11</td>
<td>245.6</td>
<td>1086.0</td>
<td>699.836</td>
<td>277.0230</td>
</tr>
<tr>
<td>Cotton</td>
<td>11</td>
<td>1275.8</td>
<td>3002.4</td>
<td>2.059E3</td>
<td>621.2462</td>
</tr>
<tr>
<td>Valid N (listwise)</td>
<td>11</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 2:
Production of Major Crops 1950 – 2003 (‘000 Tonnes)

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>Minimum</th>
<th>Maximum</th>
<th>Mean</th>
<th>Std. Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wheat</td>
<td>11</td>
<td>3235.8</td>
<td>19235.0</td>
<td>9.954E3</td>
<td>5923.0501</td>
</tr>
<tr>
<td>Rice</td>
<td>11</td>
<td>837.2</td>
<td>4487.0</td>
<td>2.544E3</td>
<td>1359.4506</td>
</tr>
<tr>
<td>Sugarcane</td>
<td>11</td>
<td>7192.6</td>
<td>52049.0</td>
<td>2.751E4</td>
<td>15153.0149</td>
</tr>
<tr>
<td>Cotton</td>
<td>9</td>
<td>1995.8</td>
<td>10211.0</td>
<td>5.964E3</td>
<td>3367.2543</td>
</tr>
<tr>
<td>Valid N (listwise)</td>
<td>9</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 3:  
Yield of Major Agricultural Crops 1950 – 2003 (Kg/Hectare)

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>Minimum</th>
<th>Maximum</th>
<th>Mean</th>
<th>Std. Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wheat</td>
<td>11</td>
<td>776.6</td>
<td>2384.0</td>
<td>1.452E3</td>
<td>578.4845</td>
</tr>
<tr>
<td>Rice</td>
<td>11</td>
<td>846.8</td>
<td>2012.0</td>
<td>1.465E3</td>
<td>403.6192</td>
</tr>
<tr>
<td>Sugarcane</td>
<td>11</td>
<td>28.2</td>
<td>47927.0</td>
<td>3.518E4</td>
<td>13007.2722</td>
</tr>
<tr>
<td>Maize</td>
<td>7</td>
<td>1160.0</td>
<td>1812.0</td>
<td>1.413E3</td>
<td>236.8086</td>
</tr>
<tr>
<td>Grain</td>
<td>7</td>
<td>452.0</td>
<td>615.0</td>
<td>528.714</td>
<td>65.3930</td>
</tr>
<tr>
<td>Cotton</td>
<td>11</td>
<td>212.0</td>
<td>621.0</td>
<td>386.945</td>
<td>159.3251</td>
</tr>
<tr>
<td>Valid N  (listwise)</td>
<td>7</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 4:  
Per Capita Availability of main Food Items 1986 – 2002 (Percentage)

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>Minimum</th>
<th>Maximum</th>
<th>Mean</th>
<th>Std. Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wheat</td>
<td>5</td>
<td>112.1</td>
<td>131.4</td>
<td>117.946</td>
<td>7.9728</td>
</tr>
<tr>
<td>Rice</td>
<td>5</td>
<td>10.6</td>
<td>20.8</td>
<td>16.376</td>
<td>4.2983</td>
</tr>
<tr>
<td>Meat</td>
<td>5</td>
<td>12.2</td>
<td>16.5</td>
<td>14.248</td>
<td>1.5557</td>
</tr>
<tr>
<td>Milk</td>
<td>5</td>
<td>56.9</td>
<td>83.1</td>
<td>69.880</td>
<td>12.2139</td>
</tr>
<tr>
<td>Valid N  (listwise)</td>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

PROBLEMS OF AGRICULTURE

- The intensive agriculture practiced by many farmers demand that problems of water scarcity, water and wind, erosion, loss of organic matter from the soil be addressed.
- Currently the country has been confronted with problems related to surplus production of some food commodities and the increased supplies are difficult to procure.
- The support [rice policies formulated to safeguard the interest of the producers and to stabilise the prices are not functioning well.
- The import of commodities like tea, edible oil and milk powder is a significant burden on foreign exchange reserves.
- The arid and semi arid rangelands that cover much of the country are degraded with chronic overgrazing and poor maintenance practices are responsible for productivity losses up to 40%.
- Marine resources of the coastal areas are also under threat. Untreated urban sewage, mangrove cutting affect partially coastal areas.
- Fresh water resources are also becoming increasingly polluted.
- Soil erosion has caused crop yield to dwindle.
- Pesticide consumption is increasingly alarming in Pakistan. Consumption of pesticides has increased from 665 tonnes in 1980 to 47,592 tonnes in 2003
- Farmers in Sindh Province in particular are adversely affected by the water crises since they are at the tail end of the canal system and they cannot use the ground-water.
- Irrigation has not been an unmitigated blessing. Canal water seeps through porous, sandy soil of the Indus basin, this poor drainage have over the years, raised the sub ground water and salt content. This has resulted in water logging and salinity over vast tracts of land.
- The weaknesses both in structure as well as methodology for technology transfer for the small farmers will lead to failure in our extension system.
- The low and stagnating intensity of research investment in Pakistan is worrisome. Compared to other industrialised countries our investment is very low.
- The existing animal quarantine facilities in the country need up gradation and capacity building in order to fulfil international quality control requirements

SUMMARY, CONCLUSIONS AND RECOMMENDATIONS

SUMMARY
The objectives of this paper were to document and examine the performance of Pakistan’s agriculture in various sectors of the economy. Identify the problems that affect growth and performance and recommend strategies for sustainability of agricultural sector in the next decade.

CONCLUSIONS
The major conclusions drawn from this paper were:

i) Total production shows that 1950 – 1955 and 2002 – 2003 for all major as well as minor crops there has been substantial increase in almost in total output over the years.

ii) In 1999 – 2000 both total production of wheat and rice fell sharply in three years as did production of all food grains.

iii) The implementation of Green Revolution created employment opportunities in crop production sub-sector.

iv) On the issue of land reforms, empirical evidence in Pakistan seems to conclusively establish that net incomes of tenants have been increased faster than those of landlords.

v) From the impact of mechanisation on income distribution, it seems that only big farmers have used the tractor-tubewell technology.

vi) Pakistan faces serious problems of depletion of its forest reserves, so far. 39,000 hectares of forest reserve is being cleared every year. If deforestation continues Pakistan will lose most of its forest within the next 40 years.
vii) During the year 2004 – 2005, a total of 90.225 million tonnes of fish and fishery products were exported, earning Rs. 7.6 billion for the country.

viii) Rural population have benefited from livestock production, household holding 2 – 3 cattle/buffaloes and 5 – 6 sheep/goats per family contributing 30 – 40% of their income.

ix) The impact of commodities like tea, edible oil, milk production is a significant burden on foreign exchange reserve.

x) Past rural development programmes were successful in terms of results. The present rural development programmes are still being implemented; we are yet to evaluate their success or failure.

RECOMMENDATION

The following measures are recommended to improve and accelerate agricultural sector for better performance:-

i) There is a need to reorganise the management of agriculture aiming at reducing the number of agencies involved and establishing effective linkages between major components of the system.

ii) Integrate the provision of inputs and supplies to the farmers possibly through a one window operation.

iii) Agriculture in Pakistan needs transformation from a traditional way of life to a modern industry to bring about a major economic uplift in rural areas.

iv) Strong emphasis need to be given during policy formulation to ensure that those responsible for agricultural production live in rural areas close to their farms.

v) Agricultural extension service should be fully equipped with the information on most appropriate land utilisation systems for different areas and relevant production technologies and should widely disseminate this information.

vi) There is an urgent need to upgrade the whole technology system for agricultural development which consists of education, research and extension.

vii) Private sector services for special operations such as rice production, transplantation, crop protection and mechanised harvesting are still not properly organised. There is need to organise them properly so as to provide the required services to farmers.

viii) A key component of government strategy must be to reshape investment and public expenditure on agriculture, spending must be focused on the provision of public goods and correction of market failures.

ix) Policies that affect incentives distort both input and output markets and results in a sub-optimal allocation of resources.

x) Changes in price policy, trade policy and fiscal policy are needed if agriculture in Pakistan is to continue to grow.

xi) The government should consider reducing its level of intervention in agricultural input markets.

xii) Rather than striving for self sufficiency in industrial output, Pakistan should direct its resources to their efficient uses and exploit agriculture’s comparative advantage to finance import requirements.
REFERENCES


A QUANTUM INSPIRED SPARSE DISTRIBUTED MEMORY (QI-SDM) MODEL FOR CLEAN PATTERNS RETRIEVAL

Sagheer Abbas*, Khalil Ahmed and Umer Farooq
School of Computer Science, National College of Business Administration and Economics, Lahore, Pakistan
*Corresponding Author: sagheer@ncbae.edu.pk

ABSTRACT

Kanerva proposed a human long-term, associative and content-addressable memory model. His theory unlocks the possibility of human-like storing and retrieving proficiency in machines. Our ominous traditional approaches are still incapable of integrating such properties in computers as it needs an ample processing power for retrieval and a huge memory for storage. It is prerequisite of time to explore other domains that are competent to overcome these complications for the simulation of memory that is close to human’s memory. Quantum Computing is the most promising approach that provides a massive parallelism as it has encouraging features of superposition, entanglement and coherence. This paper proposed QI-SDM model that simulates human like regaining competency in computers with a small cue from the memory system.

KEYWORDS

Content-addressable; traditional approaches; Quantum Computing; superposition; entanglement.

1. INTRODUCTION

Many researchers and the initiators of AI gave a strict assessment that artificial intelligence and cognitive sciences are lacking in evolvement since last few decades. Researchers should re-think and understand the things that make humans exceptional, improve computational models based on these properties, employing them in a real environment so they are able to classify and overcome the gaps by humanizing their models according to their needs.

It is often observed that our current Artificial Intelligence tools and approaches are insufficient to offer the required framework to exercise machines with human like properties. Artificial Intelligence is mostly based on the working of our brain and its working is still under graspable. Therefore, it is our requirement to grip the principles of human brain working, frameworks to be studied and tested so that the imperfections can be uncovered, leading towards more enriched and perfect models (Mateus Mendes, 2007). In the memory management systems, agents are relying on limited amounts of memory and heavy processing. However, human brain works on the other way: limited processing and huge amount of memory to store images in a sequence that are used for future predictions. It is observed that agents should rely on intelligent systems that are capable of same human like aptitudes. To simulate human like capabilities in machines is
a challenging task but it would be very useful if reproduction of these abilities in machines is successful.

Sparse distributed memory is a stimulating form of associative memory as it is widespread in both psychology and computer sciences and a plausible mathematical model of human long-term and content-addressable memory. Many researchers tried to simulate human’s similar SDM using small cue with clean pattern retrieval but they just improved the performance of original SDM. Recently, the storing and retrieving capability of modified SDM and using the Rank order SDM model have been studied profoundly and founded that these models are still insufficient to provide human’s identical storing and retrieving efficiency as it was advised by Kanerva.

The arrangement of QI-SDM paper follows as: section II defines the brief overview of the Sparse Distributed Memory model and working of SDM is clarified in section III. In section IV, the relevant literature is explored. The proposed model was discussed in section V and its working is illuminated in section VI. At the end, conclusion and future work are highlighted in section VII.

2. SPARSE DISTRIBUTED MEMORY

In 1988, P. Kanerva proposed a content-addressable, mathematical model based on high dimensional binary vectors called Sparse distributed memory (Kanerva, Sparse Distributed Memory, 1988a). In the literature of psychological aspects, Sparse distributed memory is not taken as content-addressable by computer engineer as the contents of a location are independent during a read or write operation in the long-term memory system (Kanerva, Sparse Distributed Memory, 1990). SDM is an array of storage locations and each location can contain a binary word. Storage locations are retrieved with its contents rather than the address of the location. A small part of the word is used for the retrieval of data from the memory system, called retrieval cue.

In a human brain a very large capacity of data for an extended period even sometimes for a life cycle. Humans can remember mobile phone numbers, pictures and N.I.C no’s for a long time through repetition exercise. This data is stored in the human’s long term memory and is retrieved when an input data is presented to the memory system that is plentifully close to the stored pattern (Hongying Meng, 2009).

SDM consists of two arrays. One for the location addresses and other holds the contents of the data. The location addresses are used as the reference addresses of the data vector. In the SDM model, one vector can be used synchronously for the address and the data vector. Kanerva also mentioned that the no of addresses are less as compared with the no of data space as a single address may point to many addressable spaces. The actual content space is called the hard locations. When a data is being written in the content space, it is placed in all close hard locations of the memory. During a read process from the memory system with a cue, aggregate value of all activated hard locations is obtained. Reading is not always hundred percent successful (Anwar, 1999). The reference address is used for the pooling of the hard locations where data is stored or retrieved. In traditional computers the reference address only activates a single memory location but in SDM it will activate all the hard locations within the activation radius. For the activation of the hard locations, hamming distance is computed between the reference
address and physical address. If the computed distance is less than the triggering radius, the location is added up to the activated locations else skipped. Hamming distance is actually the difference among two binary numbers \( dh(x, y) = \sum (|xi − yi|) \).

3. WORKING OF SDM

SDM working resembles beautifully to how human and animals perceive signals at two or more different times of the same location, and yet the source of the signal is identifiable as explicit place or location. If a human is given a picture to recognize, it does a few transactions with small computation at 200 MHz. On the other hand, traditional computer does thousands of transactions even with Sandy Bridge technology to recognize a single picture. In traditional machines a look up table (RAM) is needed for the information storage and it provides inflexibility due to its address-oriented practice. It is not possible to retrieve information from a partial or the degraded inputs as it requires exact address of the memory location.

In Fig: 1, let the radius is set to 3, and the input address is 00110011. Only the first and the third locations of the physical storage are activated, for their difference is less than three or equals to three. All other locations are neglected that have radius greater than 3 for storage and retrieval process in the memory system. SDM working is based on three registers: One for data input, other for address of data and third for the data out. In the above example, data input contains 01100111, address register contains 00110011 and the output register is returning 01101111 as the retrieval from the memory system with cue.

4. LITERATURE REVIEW

In the modified SDM model, several changes were made to the original SDM Model that enhanced the performance comprehensively. It significantly increases the recall performance from the memory system with a cue. Moreover, two different methods are used to implement the content matrix, in order to avoid from the complexity of implementing hardware storage. The tri-state technique fulfills the storage requirement expressively with high performance as compared to other practices. This model is able to recall clean pattern from the noisy input as it uses tri-state logic for storing and retrieving data (Kofi Appiah, June 2009).

SDMSCue shows better performance as compared to the original SDM as the traditional SDM was not able to handle small cues. It was a major drawback of the original SDM to recover associations comparatively with small cues. In original SDM, an
acceptable large portion of the previously written data must be presented to the memory model as an address for the retrieval purposes. The SDMSCue is the enhanced version of the original SDM. The capability of SDMSCue provides better performance, overcoming the input length restriction in the traditional SDM model. The recall results obtained based on the associations for SDMSCue are relatively better to those of original SDM (Ashraf Anwar, 2004).

In 2009, Manuel et al explored that according to Kanerva’s SDM model, the hamming distance should be used to compute the resemblance between two memory items. Unfortunately, it shows a poor performance provided that the data is not random. The NBC with the hamming distance shows the worst performance. If bits are grouped as bytes and an arithmetic distance is used instead of hamming distance, performance is improved exceptionally but it fades some characteristics of the original model that is based on the properties of a binary space (Manuel Crisostomo, 2009). In 2003, Joy Bose adopted a different method for simulating human like capacities in machines with the help of neural network but slightly improved the performance of the original SDM (Bose, 2003). Many researchers have implemented SDM after some minor or major changes in the original model to meet the desired results that were advised by Kanerva, but not succeeded to simulate human’s like memory practice except some positive results inform of enriched performance.

SDM has some limitations. It provides low storage that can be as low as 0.1 bits per bit of ordinary memory and an additional processing is a prerequisite to access the memory, when it is exhibited in software. (Ashraf Anwar, 2004), (A. Paulo Coimbra, 2007) Classical SDM is causing to create human-like retrieval capability in machines. Our current artificial intelligence architectures are the main obstructive part for simulating human-like retrieval with a cue from the memory system. Classical computers have very complex architectures and computations, so it is required to explore extreme better ways for storing and retrieving data in the memory system. Quantum mechanics is a promising approach as it has possibly motivating properties of entanglement, interference and superposition of states. Quantum-Inspired SDM will be very similar to look like human’s storage and retrieval proficiency in the memory system.

In 1982, Feyman explored the idea of a computational device based on quantum mechanics. He was the first who showed how a quantum system could be used to simulate classical computations and explained how such a machine would be capable to act like a simulator for quantum physics. (Richard, 1982) Feynman also answered to this question by producing an intangible model while simulating quantum computer on traditional computer (Richard, 1985). Later, in 1985, Deutsch appreciated that Feynman’s assertion can ultimately lead to a quantum machine. He published a theoretical paper to display principally that any physical process could be demonstrated perfectly by a quantum machine. Thus, a quantum computer would have much better proficiencies as compared to any traditional computer. After this paper, the search began to find encouraging applications for such quantum machine (Deutsch, 1985).

In 2003, Carlo emphasized in quantum pattern recognition that the amplitudes of the rotated memory state of the stored patterns that are close to the hamming distance are peaked resulting a memory pattern very similar to it. He highlighted that, in humans,
recognition process is based on our associations of memories that are related to specified cue. Moreover, he declared that quantum mechanics compromises a way-out from the inconvenience of incorporating the association power of content-addressable memories without growing the hardware complexity. Quantum mechanical entanglement deals with a natural phenomenon for refining both the storage and the retrieval capacity of associative memories with corrupted or incomplete information (Trugenberger, 2003).

In 2008, Zhou et al presented an innovative method ‘Quantum pattern recognition with probability of 100%’ that is able to recognize multi-pattern concurrently with 100% probability. It was not possible for any other pattern recognition method to accomplish this task with 100% probability (R. Zhou, 2008). The theory of quantum computation may probably be traced back to the revolutionary work of R. Feynman, who observed that the role of quantum effects would play a major part in the development of future hardware. Feynman also mentioned that such quantum devices can have considerable rewards over classical computational methods (Alexander A. Ezhov, 2005).

5. PROPOSED MODEL

The proposed model comprises of two rational divisions, classical part and the quantum part of the memory model system. To date, it is not possible to simulate Quantum Inspired SDM on quantum machines due to the unavailability of quantum machines but quiet practical to simulate on traditional computers to validate the model. Although the performance of the model will vary on both machines but it will validate our model. A peer to peer network of traditional computers is desired to simulate QI-SDM on these machines for boosting the performance of the model.

The classical part consists of the input, address matrix and the output part of the model. A classical input is accepted by the model for storing or retrieving data in the memory. If a cue is served to this QI-SDM for the retrieval initiatives, the final output will also be the classical one.

The quantum part is the prime focus of the model. It comprises of the content matrix, devising the address locations and the hard locations of the model respectively in the beginning. QI-SDM exploits superposition, entanglement, coherence and interference to achieve parallelism that is not possible in classical computer. This parallelism could lead to memory system with human like proficiencies and exponentially faster than traditional version of Sparse Distributed Memory.

Figure 2: Quantum Inspired SDM
“Data in” is the register that contains the data that going to be stored in the content matrix of Quantum Inspired SDM against the activated rows or the data that is being used to cue the data from the data matrix and consists of m bit-pattern. During or a storage process data against the activated rows is stored in a single go, shifting its state to another state. “Data out” is the final output that is gained after the convergence or divergence from QI-SDM from the memory system and it is classical output after measuring the quantum states and threshold at ‘0’.

Cue consists of the pattern of n qubits that is used as partial information to access the data from the quantum sparse distributed memory. In this model, Cue is used as a hint for retrieving data from the memory system. Address Matrix is a matrix that holds ‘M’ addresses of hard locations with n qubits patterns that are available for QI-SDM during any read or retrieve information process. These addresses are once associated with the hard locations can’t be changed, as they work as reference address only. Content Matrix is also a matrix that is the tangible space where data is stored against the triggered rows in this model. When a cue is used for storing or retrieving data from the model then rows in the content matrix are stimulated against the triggered reference addresses.

A quantum memory register of n bits can exist in any one of \(2^n\) base states, and a quantum register may exist in a superposition of those base states. In this model, two entangled qubits are taken together that’s means the quantum register can be measured to any one of four states (\(|00\rangle, |01\rangle, |10\rangle, |11\rangle\)). Measurement is a process that is used to extract classical value from a quantum system against the activated rows of the content matrix. The probability that the state is measured as basis vector \(|a\rangle\) is the square of the norm of the amplitude of the component of the distinctive state in the direction of the basis vector \(|a\rangle\). For example, given a device for measuring the polarization of photons with associated basis \(|\uparrow\rangle, |\rightarrow\rangle\), the state \(|\psi\rangle = a|\uparrow\rangle + b|\rightarrow\rangle\) is measured as \(|\uparrow\rangle\) with probability \(|a|^2\) and as \(|\rightarrow\rangle\) with probability \(|b|^2\).

Let us start with an example for measuring a state in a two-qubit system. Two entangled qubits state can be stated as \(a|00\rangle + b|01\rangle + c|10\rangle + d|11\rangle\), where a, b, c, and d are complex numbers such that \(|a|^2+|b|^2+|c|^2+|d|^2 = 1\). Let’s start; we are interested for measuring first qubit with respect to its basis states \(|0\rangle, |1\rangle\). For our convenience, it can be rewritten as follows:

\[
a|00\rangle + b|01\rangle + c|10\rangle + d|11\rangle = |0\rangle \otimes (a|0\rangle + b|1\rangle) + |1\rangle \otimes (c|0\rangle + d|1\rangle)
\]

For \(u = \sqrt{|a|^2 + |b|^2}\) and \(v = \sqrt{|c|^2 + |d|^2}\), Once the state has been rewritten as above, as a tensor product of the qubit being measured and a second vector of unit length, the probabilistic result of a measurement is easy to read off. Measurement of the first bit will with probability \(u^2 = |a|^2 + |b|^2\) return \(|0\rangle\), measuring the second bit works similarly.

Measurement provides another way to consider about entangled particles. Qubits are not entangled if the measurement of one has no effect on the other. For instance, the state \(1/\sqrt{2}(|00\rangle + |01\rangle)\) is entangled, since the probability that the first bit is measured to be \(|0\rangle\) is 1/2 if the second bit has not been measured. If the second bit of the given state had been measured, the probability that the first bit is measured as \(|0\rangle\) is either 1 or 0, subject
to whether the second bit was measured as $|0\rangle$ or $|1\rangle$ respectively. Any measurement of the first bit will yield $|0\rangle$ regardless of whether the second bit was measured. Similarly, the second bit has a fifty-fifty chance of being measured as $|0\rangle$ irrespective of whether the first bit was measured or not. Note that entanglement, in the sense that measurement of one particle has an effect on measurements of another qubit, indicates that entangled states as states cannot be written as a tensor product of individual states. Two qubits based QI-SDM model has two basis states which can be prominent by measurement in the computational basis, giving a classical value. A superposition state comprises amplitudes for $|0\rangle$ and $|1\rangle$ at the same time. For instance, in the superposition state, $|0\rangle$ and $|1\rangle$ has equal amplitudes for each basis state, meaning that there is a 50% probability of measuring the qubit in $|0\rangle$ and a 50% probability in $|1\rangle$.

Superposition’s of quantum states are the basis of the interference that drives a quantum computer; quantum algorithms try to manipulate the amplitude and phase of various states so that predicted states (the answers to the problem being solved) have a high probability of being measured while the undesirable states (the non-answers to the problem being solved) have a low probability of being measured. Quantum superposition and entanglement generate an extremely enhanced computing power. Where a 2-bit register in an conventional computer can hold only one of four binary configurations (00, 01, 10, or 11) at any given time, on the other hand a 2-qubit register in a quantum machine can store all four numbers simultaneously, because each qubit represents two values. If the numbers of qubits are added, the enriched capacity is extended enormously. In quantum machines, quantum states are manipulated instead of classical states. A basis for a 2-qubit system is $\{|00\rangle, |01\rangle, |10\rangle, |11\rangle\}$.

6. WORKING OF QI-SDM

Considering a classical computer that works on a 3 bit register, at any given time, the bits in the register are in a definite state, such as “101”. On the other hand, in a quantum computer the qubits can be in a superposition of all the naturally acceptable states. In fact, the quantum register is defined by a wave function:

$$|\psi\rangle = s|000\rangle + t|001\rangle + u|010\rangle + v|011\rangle + w|100\rangle + x|101\rangle + y|110\rangle + z|111\rangle$$

Let $|\text{Reg}\rangle$ be the input register to a quantum system that grips all possible states of a three qubits register. A quantum register with three qubits, can obtain 8 superposition states ($2^3 = 8$) simultaneously. Classically the value stored in the content matrix of the 3-bit register will be 0 or 1 at a time. Conversely, the value stored in the content matrix of the QI-SDM is in the superposition state. That’s means all eight states can be stored in superposition state using the wave function.

$$|\psi\rangle = \begin{bmatrix}
|000\rangle & |001\rangle & |010\rangle & |011\rangle & |100\rangle & |101\rangle & |110\rangle & |111\rangle
\end{bmatrix}$$

Qubits are proficiently arranged in a way that are capable of swapping from state one to another where s, t, u,…z. coefficients are complex numbers that are depending on the amplitude squared , that are the probabilities to degree in each state. For example $|s|^2$ is the probability measurement for the register in the state 000. Practically and critically, the
phases of these numbers interfere with one another, is the leading stuff of quantum algorithm.

It is conventional that a huge number of classical bits are required even to estimate the complex numbers of some quantum state rises exponentially with the number of qubits. For a simulation of a three hundred qubits quantum register, 1090 traditional registers are needed that are larger than the total atoms that are noticeable in the universe.

In figure 3, let’s start with the input data ‘01101111’ that is presented to QI-SDM. It will stimulate the rows against the address matrix, having radius equal or lower as compared to the input data. The input is stored in a single step, from state one to the other, due to the superposition states of the qubits that are devised to content matrix. During a read operation, against the activated rows, classical value of the content matrix is extracted, summed and averaged. Finally, the output values are converged at its threshold level at 0.5

![Figure 3: Working Of QI-Sparse Distributed Memory](image)

7. CONCLUSION

QI-SDM model improves the performance of SDM significantly during a read or write operation for extracting clean pattern with corrupted or incomplete cue due to the fascinating properties of quantum mechanics. It also reduces the storage requirement significantly due to its ability for storing multiple states in superposition states. This model is pretty establish-able in agents and robots and ideal for hardware implementation. Our future work will focus on the implementation of powerful and parallel QI-SDM as a clean pattern extraction and recognition machine.

REFERENCES

EMOTIONAL STATE TRANSITION MODELLING USING ARTIFICIAL NEURAL NETWORKS

Atifa Athar and Khalil Ahmed
School of Computer Science, National College of Business Administration and Economics, Lahore, Pakistan
Email: atifaathar@yahoo.com

ABSTRACT

Emotional consciousness is identified as an important aspect of human personality and provides the basis for rationality. Emotion modeling is a key area under consideration by the researchers in the recent years. The difficult aspect of emotion modeling is to develop a framework to assess the impact of environmental changes on different emotional states. By simulating the human mind processing, the aim of this paper is to present computational process model using neural networks to address the emotion transition from one state to another.

By using neural networking techniques this model has the potential to keep track of nonlinear relationships between different emotional states and secondly it provides the foundation for learning emotion generation.

1. INTRODUCTION

The intelligence phenomenon in humans is attributed by emotions. The emotions are realized by human mind when it senses an external stimulus from the environment. From the neuropsychological perspective it is known that emotions do not mix with each other and human mind cannot be confused while experiencing more than one emotion. Each emotion is considered as a discrete entity and can be differentiated from other emotions on the basis of its properties. These properties mainly include the valence and intensity of an emotion.

Emotions are not constant in nature. As the valence or intensity level of the current emotion changes it converts into a new emotion. Emotional change includes shifts from one emotion to another due to the change in environment and dispersing some emotions. (Athar & Ahmed, 2008) Excited duration of an emotional state is also an important feature of change in emotional states.

Before moving to the emotional state transition process, one should have an account of how many emotions does exist. Each emotion is not a single affective state but a family of related states. (Dalgleish & Power, 1999) Hence there is the concept of emotional families discussed by different theorists. Each emotional family consists of emotions which has some common properties. These common properties help us to differentiate between different families while each emotion can be identified by the properties which are unique to it in the same family. Ekman (Ekman, 1999) explained these similarities as themes and differences as variations. According to him “The theme is
composed of the characteristics unique to that family; the variations on that theme are the product of individual differences, and differences in the specific occasion in which an emotion occurs. The themes are the product of evolution, while the variations reflect learning.”

The emotional state transition process of mind explains that although each emotion has distinct properties but it can evolve due to frequent encounters with the same stimuli whether internal or received from the environment. Secondly the emotional change can arise due to the excitation of any related emotional object that is already a part of emotional experience and resides in memory.

Emotions are not an isolated arrangement of mind but as Lazarus (Lazarus, 1991) explained that emotions represent transactions with the environment. These transactions can be measured in terms of the degree of relationship between an emotion and its related environmental cue. This paper presents a classification of different emotional states and transition due to the internal or external stimulus.

De sousa (De Sousa, 1987) suggests that learning emotions is like learning a language. As each language is based on some syntax and semantics, the semantics of emotions are based on situations which cause their generation. Through the evolutionary process the emotional semantics have been learned by human mind and became the part of experience. This experience grows and populates the human memory as the new transactions takes place.

It is a fact that the field of emotion modeling is under consideration from the last twenty years but the models presented were not very appropriate to accommodate the computational framework of emotions which is based on the cognitive neuroscience approach. Despite some contributions by earlier scientists there have been small attempts to use neural networks for emotion modeling. However, most of these models operate at a high level of conceptual abstraction and rarely include the underlying neural architecture. (Sander, Grandjean, & Scherer, 2005)

The functionality of biological model of human mind turns out to be the motivation for the construction of emotional state transition model using an emotion oriented artificial neural network. ANM provide a high level of specificity to devise a linkage between theoretical suppositions and neural dynamics.

The resemblance between human mind and neural nets regarding learning ability is another key point to focus. Using ANN modeling techniques the emotional experience can be stored and updated in the artificial neural repository in the form of patterns.

Previously presented models by different scientists, were based on two dimensional neural spaces by considering the fundamental aspects of different emotions. There is another dimension regarding evolution is observed and therefore the state transition model presented here is based on three dimensional neural spaces

2. CLASSIFICATION OF EMOTIONAL STATES

Emotion terms are discrete, but emotional states form a continuum. Discrete terms need to be embedded in continuous representations that at least permit interpolation, and
that ideally allow for the kind of shaded judgment that people can achieve (e.g. by qualifying emotion terms). (Cowie, Douglas-Cowie, Apolloni, Taylor, & Romano, 1999).

An emotional state is based on some conditions. These conditions are formed by checking the level of homeostasis (search for balance) and hetero-stasis of an emotion.

Emotional states can be classified on the basis of commonalities and variations, as follows

2.1 Primary Emotional States
Each emotion is unique and the primary emotions can be distinguished in terms of their properties i.e. intensity and valence. So fear, anger, sadness and disgust are negative emotions while happiness, satisfaction, contentment and relief are positive ones at different level across the intensity axis.

Primary emotions can also be separated in course of their evolutionary nature. Tooby & Cosmides (Tooby, 1990) tell us that emotions impose on the present world an interpretative landscape derived from the covariant structure of the past ...”. Emotions, they say, deal with recurrent "... adaptive situations ... fighting, falling in love, escaping predators,... , infidelity, and so on, each [of which] recurred innumerable times in evolutionary history ...” (pp. 407-408). The crucial element which distinguishes the emotions: our appraisal of a current event is influenced by our ancestral past. (Dalgleish & Power, 1999) A third dimension of emotions could be observed other than intensity and valence, named as stance to accommodate its evolution from the past experience.

The primary emotional states are characterized by an intrinsic response to the situation or event. These states such as being startled or terrified are typically triggered by patterns in the early sensory input and detected by a global alarm system. (Allen, 2000).

2.2 Secondary Emotional States
The association between primary emotions caters the possibility of complex or secondary emotions. For example, smugness might be considered to be a blend of the two elemental emotions, happiness and contempt. (Dalgleish & Power, 1999)

The primary emotions are related to each other to form a full scale emotional experience just like we can make new colors from existing primary colors. (Athar & Ahmed, 2008) The emotional experience utilizes the secondary emotions to improve the performance in terms of rationality.

Fig. 1: Mapping of emotional categories to arousal, valence and stance dimensions (Breazeal, 2003)
Secondary emotional states such as being fearful, worried, or relieved could be the result of performance assessments in terms of risks, progress and directed success.

### 2.3 Tertiary Emotional States

The tertiary emotions are highly complex and multifaceted states which emerge from the primary or secondary emotions. For example: Nostalgia is a combination of happiness, sadness and desire. (Rue., 2005) It is conceived that tertiary states are hard to enter into. Tertiary emotions correspond to secondary emotions which reduce self-control. (Allen, 2000)

Tertiary emotional states are characterized by powerful control and complex but rich semantics.

We can plot the emotional states of a human mind tackling any situation. For example a student’s emotional states while working with an assignment using the State-Transition Diagram. It is clear from the diagram below that how triggers can lead to their changing states of emotions from the negative to positive and vice versa. i.e. a concerned student can get annoyed if do not get the required material for assignment.

This diagram also depicts that change in emotion is not just due to the change in its valence but it is also due to the change in the intensity level. A concerned student could be confident if gets some related material. i.e. a positive emotion could be changed into another positive emotion have high intensity.

An emotional state can be represented as an integration of two or more emotions having same valence but different intensity levels like a student can be happy and relieved. This integration provides an example for complex emotional state i.e. secondary or tertiary emotional states.

### 3. EMOTIONAL STATE-TRANSITION MODEL

Simple classification of emotional states would not be much beneficial unless it would be integrated with a practical CS approach. This paper presents a computational model to cater emotional state-transition process using artificial neural networks as it is a nonlinear modeling technique for classification, transformation, association and process control. (Zhecheva, Malinov, & Sha, 2005).

It is known from the literature review that many authors agree that there are mainly two dimensions of emotions to be measured. i.e.
evaluation and activation. Evaluation is measured in terms of valence (how pleasant or unpleasant an emotion is) and activation in terms of intensity (level of arousal).

According to OCC model intensity can be measured as a scalar but it is represented as a pair of positive and corresponding negative emotions by poel. (Poel, Akker, Nijholt, & and Kesteren, 2002) For example love and hate can be measured by the same scalar value for intensity. Further a positive intensity value shows pleasant emotion (love) and the same negative intensity value corresponds to the unpleasant emotion (hate). But the drawback of this model is that it cannot generate positive and negative emotions of same pair at a time.

Cynthia Breazeal (Breazeal, 2003) considered an emotion as a specific set of computational processes which are active in the neural network system. So there are two possibilities for each emotion i.e. to be active or passive. Each emotion whether active or passive, has an associated vector regarding its properties which consists of its intensity, valence and stance \([I, V, S]\). The values for this vector determine the current state of an emotion.

The effectiveness of each environmental cue can be calculated by the vector having same three properties i.e. intensity, valence and stance. Each of these has some associated values which contribute to maintain overall effectiveness of the environmental cue in the following way. The intensity specifies that how much arousing this event is to the upcoming emotional state. Positive value for I shows the high arousal of stimuli whereas negative values show the low arousal stimuli. The V shows how pleasant or unpleasant is the stimuli for the emotional network. Where a positive value corresponds to the pleasantness and negative value corresponds to unpleasantness. The label S specifies the stance of the stimuli for reoccurrence. Positive values for S allows for the evolution whereas negative values specifies retreatment.

The property vector of a new emotional state depends on some elements that cause the current emotional state to be turned into the new one. These elements include the parameters of current emotion and the parameters of the environment or event cue which may be internal or external. The emotional state transition process model demonstrates the important correlation of all these processing parameters.

The architecture of the feed forward neural network for the simulation and prediction of the new emotion is organized into three layers, an input layer: consisting of two groups, one is current emotional state and other is an event cue, an output layer: consisting of the parameters of new emotional state and a hidden layer: which would optimize these processing elements with excitatory and inhibitory neurons. The number of neurons in I/O layers depends on the I/O parameters while different number of neurons

<table>
<thead>
<tr>
<th>Pos. Emotion Types</th>
<th>Neg. Emotion Types</th>
</tr>
</thead>
<tbody>
<tr>
<td>Joy</td>
<td>Distress</td>
</tr>
<tr>
<td>Hope</td>
<td>Fear</td>
</tr>
<tr>
<td>Satisfaction</td>
<td>Fear-confirmed</td>
</tr>
<tr>
<td>Relief</td>
<td>Disappointment</td>
</tr>
<tr>
<td>Pride</td>
<td>Shame</td>
</tr>
<tr>
<td>Admiration</td>
<td>Reproach</td>
</tr>
<tr>
<td>Happy</td>
<td>Resentment</td>
</tr>
<tr>
<td>Gloatting</td>
<td>Pity</td>
</tr>
<tr>
<td>Love</td>
<td>Hate</td>
</tr>
</tbody>
</table>

Fig. 3: The positive (Pos.) and negative (Neg.) emotion types according to OCC model
for the hidden layer are attempted to reach the best architecture to produce optimum solution.

As ANN function for quantitative parameters so input and output parameters [I,V,S] are presented in numeric format. Both input and output values are normalized between -1 and 1.

\[
x_{\text{norm}} = 2 \left( \frac{x - x_{\text{min}}}{x_{\text{max}} - x_{\text{min}}} \right) - 1
\]

Where \(x_{\text{norm}}\) is the normalized value of a certain parameter, \(x\) is the measured value for this parameter, \(x_{\text{min}}\) and \(x_{\text{max}}\) the lowest and highest values in the database for this parameter, respectively. (Zhecheva, Malinov, & Sha, 2005).

The following steps are performed in a complete iteration of the neural network. The parameters of current emotion and event cue are presented to the input layer using six neurons. Then weights are assigned to each input and presented to the hidden layer. The impact of each dimension of event cue on each parameter of current emotion is computed individually in the hidden layer which consists of nine neurons initially. Current emotions are updated and the properties of new emotions are determined and presented by the output layer which consists of one neuron.

Learning is an important aspect of ANN modeling technique. The learning (training) is achieved by adjusting the weights of connections between neurons in each layer. These weights are updated by using back propagation algorithm for this state transition process model. The training of this model is achieved through the following steps.

The training starts with six neurons in the input layer, one neuron in the output layer and nine neurons in the hidden layer. The purpose of hidden layer is to compute the
correlation of all parameters and is to determine that which parameter contributes more to the new emotion. The hidden layer helps us to filter out each contribution individually and independently. For example a great negative value for valence does not only contribute to “fear” but to “anger” and “disgust” too.

In the second step the number of neurons in the hidden layer is reduced from nine to six as all the combinational contributions of the incoming parameters cannot be accepted to calculate the net parameters of the new emotion. For example if current emotion is “tired” (low intensity state) and a threat appeared (high intensity and strong negative valence). This will result the neutral arousal and slightly negative valence which is insufficient to evoke fear.

In the third and final step the neurons in the hidden layer are again reduced from six to three.

4. CONCLUSION

An artificial neural network model for the simulation and prediction of emotional states has been designed in this paper. Initially this emotional state transition model depicts the parameters of an affective state of event cue and then their influence on the emotional parameters is calculated.

In comparison to the neuropsychological aspects of emotional states of the human mind this model can be used to predict the new expected emotional state.

REFERENCES

ROLE OF EMOTION ORIENTED AGENTS IN INFORMATION MEDIA FOR EFFECTIVE COMMUNICATION

Atifa Athar and Khalil Ahmed
School of Computer Science, National College of Business Administration and Economics, Lahore, Pakistan
Email: atifaathar@yahoo.com

ABSTRACT

The role of emotions is observant in effective and focused communication that is one of the social functions of human life. Keeping in view that emotion expressions and recognition are specific to the social context these are required to achieve information fidelity in operational media communication. Agents are becoming human companions in a variety of social aspects and their role as our future social partners suggests that they are required to be equipped with emotions in order to make media communication adequate and acceptable.

In this research we are arguing that emotion oriented agents are plausible in information media to realize natural communication.

1. INTRODUCTION

Emotions are social things (Gillian Bendelow, 1998) and frequently used in communication and are considered as an affective component of human mind that subjects to the experience of varying circumstances. These are largely considered in terms of helping an agent to adapt to limitations, to manage social behavior, and to communicate with others. (Robin Roberson Murphy, 2002)

Agents as our future social partners need to be facilitated with emotions. If tomorrow’s robots are going to be the part of our world, they should possess some emotional behavior that allows them to communicate and respond in styles people can understand. (Hudlicka E, 2009) Especially while communicating information through media, emotions can provide control and derive the behavior of an agent.

The role of emotions in information media processes has a potential of holding broad collection of communicative contexts, such as health, politics, advertising, entertainment, educational programming, online communication, etc. (Nabi & Wirth, 2008) These days’ information media is getting diverse and customized and it acts as a producer of social change. (Scott A Reid, 2004) To deal with the state of affairs and their belongings in communicative stage of information media emotions are unavoidable. Emotional information exchange is a key feature in media communication among humans to get connected with the audiences.

Different researchers have highlighted the effects of media and emotions on each other but the concept of employing emotion oriented agents in information media is still infancy and arguable. The impetus for employing emotion oriented agents in information
media is that it would enable the agents to adapt to the uncertain information either risky or beneficial, with causal reasoning and communicate it to the audience in a reliable manner. This investigative research would serve to identify that how emotionally intelligent cohorts would be capable of maintaining long term and trustable communication with the audience of information media.

2. COMPONENTS OF EMOTIONS

According to Scherer, emotions are considered as complex interactive things including affective, cognitive, conative, physiological, and motivational components. (Scherer, 1984) Nabi explained the function of each of these components in detail. The affective component corresponds to the situational contexts that provide the emotional experience which is connected to the level of arousal, pleasantness and satisfaction. The cognitive component is for the perception development and evaluation of the environment. The conative component is associated with the expression of emotions felt while physiological component is to mediate emotional reactions. Lastly motivation component corresponds to the behavior and intentions. (Nabi & Wirth, 2008)

Research on media and emotions is emphasized on either the discrete or dimensional theories of emotions. The dimensional theory consists of emotion characterization by arousal, valence and dominance while theory of discrete emotions considers individual emotional states. Both theories provide a number of ways to model emotion for media processes.

3. DIMENSIONS OF SOCIALITY AND MEDIA: PHYSICAL VS. SOCIAL

The emotion oriented communication means the recognition, expression and sharing of emotions between two or more individuals. The comparison of implicit and explicit communication identifies the worth of emotional contextual difference. The different dimensions of the social presence of individuals or audience of communication are also important to consider. (Daantje Derks, 2008)

To address the two dimensions of sociality i.e. physical and social in media communication the use of emotion oriented agents is suitable. The physical dimension is about the actual co-presence of being together at the same place with bodily contact and visibility of expression while social dimension refers to the less visibility and unawareness of co-presence.

It is observed from the studies by different researchers that physical presence of an emotional agent is more acceptable in comparison to the virtual agents or a graphical view in different communication scenarios. (Pereira, 2008) Attaining likeness in emotional expression of humans and agents is a key to the successful sharing of information on media.

The differentiation and impact of empathy for emotional characters and emotional responses of the audience to the agent’s emotions are the parameters to form the connectedness between them in media. The indulgence of audience to the emotional protagonist creates persuasion in them towards the information communicated.
4. INFLUENTIAL FACETS OF EMOTION ORIENTED AGENTS ON INFORMATION MEDIA

Since Information media is a complete system having different components, and in order to make this system grow and perform it is required to enhance the communication between its components. Considering artificial agents as our future social partners and component of information media, their communication with other components is surely essential. Since emotions are identified as careers of information, employing emotion oriented agents in media is adequate for effective communication in view of the following aspects.

4.1 Connectedness
Conveying true emotional states serves both, the sociality and media. (Gratch, 2008) The prospect of connectedness between emotional agent and media audience depends upon the level of sensitivity and shared experiences. While communicating information on media it is important to identify, that what to say and how to say with respect to knowledge acquired before. To achieve this goal embodying emotions in agents serves the purpose with reasoning and prioritizing capabilities.

Social intelligence in the form of emotional expression has an impact on effective communication and is a way to build a persistent relationship between emotion oriented agents and audiences. It is also a critical aspect of computer based media interaction. People naturally respond to all online interactions as a social experience.

4.2 Persuasion
Audience can be educated, stimulated and convinced by the emotion oriented media partners. The emotion oriented agents in media communication may change the course of interaction and belief as they can express emotions appropriately, show enthusiasm and acknowledge or criticize. By personalizing the information these agents could make audience to reflect socially. (Reeves, 2000)

4.3 Unification
Emotions and social world are constitutive to each other. (Tiedens & Leach, 2004) The sharing of emotions between audience and emotional agents can lead them to unified emotional expression with the concept of convergence. This emotional convergence shows the interrelationship between media partners.

4.4 Interfacing
Emotions are significant in human interaction. They can have effect on memory, attention and feeling at any specific moment. The agent with the ability to express emotions while communication can provide cues that what is important. The emotions in communication serve to manage behavior. Along with verbal communication the emotional expression signals to audience more rapidly and in an absolute manner.

4.5 Learning
The learning and business are social aspects for humans. Learning is considered as a facet of human conversation. The learning rate could boost up with the continuation in interactions to have larger amount of personalized experience.
The role of media is substantial in learning and business promotion. Emotion oriented agents can play a social role to let audience provide an expert’s opinion on what to buy and sell and how to do it. A complete variety of feelings could be generated in the audience through accruing emotions and behaviors by employing emotional agents in information media.

4.6 Environment Modification

Emotions are considered dynamic and fractional and provide momentary understanding of the physical and social context which continuously evolves from changes in the environment, inference, and social feedback. (Gratch, 2008) Alternatively intensity of an environmental event depends upon the sharing level of the relevant emotions by the media partners.

4.7 Controlling Cooperation

The social functional view of emotions offers a tight connection between emotional processes and emotional behavior – we display what we feel – and this rapid, involuntary and authentic nature of emotional displays is, in fact, necessary for facilitating cooperation and group cohesion (Gratch, 2008) and balancing ego-emotions and co-emotions.

Achieving control over the cooperation between audience and media partners is possible with the representation of social intelligence. The keenness, confidence and having right information to deliver are the parameters of controlling the level of cooperation.

4.8 Maintaining Social Norms

Emotions can respond to, be regulated by and constitute social relationships. (Tiedens & Leach, 2004) Therefore emotion oriented agents can be used to manage the compelling or conflicting relationships among different audiences. It is a salient feature of communication that helps in maintaining social norms by creating, maintaining and dissolving social relationships among the audiences.

4.9 Recognition and Recall

Emotional events are better recalled than normal events because strong emotional content are better remembered by humans as compared to the neutral contents. According to psychology, positive emotional content is well memorized and communicated then negative emotional content and thus improves the wellbeing. In media communication an agent with expressive emotions could make human audience better memorize the information delivered and recall it latter.

4.10 Branding

The likeness in the personalities of sender and receiver makes communication very effective while unpredicted change in them forces communication to be negotiated. Emotions are considered the main constructs of personalities. (Reeves, 2000) The use of emotion oriented agents in media is a reliable way to reflect a corporate brand as they can operate consistently.

4.11 Impact of Emotion Expressivity on Well-Being Promotion in Audience

A reason that emotion oriented agents is acceptable in information media lies in their potential to address the communication partners. The employment of affective agents is
also beneficial for knowledge acquisition and learning. Considering all influential aspects of emotion oriented agents on information media creates the feeling of personalization and social caretaking in the audience. This sense of personalization leads to the development of trust and thus promotes the well-being.

Responsiveness of partners is a key to effective communication in media that needs awareness. Awareness is considered as a consequence of inference and emotions are well thought-out as inferential shortcuts. (Athar, Ahmed, Qazi, Saeed, & Saeed, 2010) That is how agents with emotional expressions could relate to the awareness and provide a route for effective and responsive communication as a social partner in the information media. The level of awareness and well-being of each communication partner could help in achieving smooth flow of information.

5. PARAMETERS FOR THE EXPRESSIVITY OF EMOTIONS

Emotional facial expressions can provide information in verbal or non-verbal communication. From the literature it is known that agents having ability to express emotions while communication can bridge the intellectual gap between human beings and machines.

Delay, duration, onset, offset and intensity are the five parameters that have been identified regarding emotional expression during communication. (Bevacqua, Pelachaud, & C. Mancini, 2004) Aiming at dragging agents having emotions in media communication it is needed to map the emotion parameters to the expression parameters.

6. CONCLUSION

This paper discusses the viability of modeling and engaging emotion oriented agents in information media with a number of interesting aspects for natural and effective communication with the audience.

The potential of emotion oriented agents with the diverse capabilities of information processing like learning, cooperation and maintaining social norms as a source to reduce disbelief and increase reliability in the audience is also debated. Feedback of audience is identified as an important factor of prosperous communication in information media and emotion oriented agents can call forth the response naturally from the audience.

In the light of this argument the characterization of emotional agents for information media as a successful career of information is suggested.

REFERENCES


PERFORMANCE EXCELLENCE THROUGH INTEGRATION OF KM PHILOSOPHY IN TQM PRACTICES

Shahid Mehmood*, Suleman Aziz Lodhi and Khalil Ahmed
School of Business Administration, National College of Business Administration and Economics, Lahore, Pakistan
*Corresponding author: shahidiub@hotmail.com

ABSTRACT

Performance is a vital subject in corporate sector; it is highly dependent not only on internal organizational functions but also equally dependent on external environmental factors. This paper presents a framework to incorporate knowledge management (KM) philosophy as a supportive segment in total quality management (TQM) practices to enhance performance excellence. The infusion of KM philosophy in TQM practices will provide multidimensional approach for performance enhancement by establishing a significant base for decision making.

KEYWORDS

Total quality management; Knowledge management; Performance excellence.

1. INTRODUCTION

In recent decades Knowledge management has gain significant popularity in business organizations. Thus, some definitions of Knowledge and Knowledge management are given by numerous scholars and researchers. Nonaka (1991) defined knowledge as acceptable belief, whereas beliefs are utilized to give explanation for self-interest. Firms must focus on organizational learning to increase knowledge in order to achieve continuous improvement in long run. (Baker and Sinkula, 1999). Knowledge has become a significant resource in the modern organizations and considered as a fundamental for gaining competitive advantage in a knowledge-based view of organizations, Grant (1996), Spender (1996) and later by Cole (1998). “Knowledge management is the process in which company’s capture their expertise by retaining it in databases, on paper, or in people’s minds – and use it in order to get maximum benefits” (Hibbard, 1997) Harrison and Leitch (2000) advised that in order to compete with the competitors organizations must continuously update their knowledge resources.

In current era the aforementioned infusion of knowledge management is being used in various sub domains of management science to equate or evaluate multi-facet effects. Total quality management is one of the critical areas which influence many other factors.

TQM is all about business management values consisting of different principles that help in continuous improvement and it is considered as the most suitable approach in sustaining efforts for organizational improvement (Lin and Ogunyemi, 1996). Importance of TQM as a source of competitive advantage is well-known around the world, and few
companies can meet the expense of ignorance of TQM (Dean and Bowen, 1994). When TQM is successfully implemented it is helpful in gaining sustainable competitive advantage (Prajogo and Sohal, 2004) Examples are available from previous research that many companies attained sustainable competitive edge by implementing the TQM practices. Thus, numerous scholars and researchers identified that those business organizations which are making efforts to strengthen their organizations must focused on TQM as the source of sustainable competitive advantage (Terziovski, 2006). Mukherjee, Lapre and Van Wassenhove (1998) recommended that the primary connection between organizational achievement and change in knowledge structures is TQM. If companies can acquire, manage and utilize organizational knowledge, they can sustain their competitive edge (Davenport and Prusak, 1998; Desouza and Evaristo, 2003).

2. LITERATURE REVIEW

KM increases the capability of a company to collect and organize the knowledge in order to improve the decision making ability and strategy formulation process (Hsu and Shen, 2005; Ooi et al., 2009). For De Jarnett (1999), KM consisting of knowledge creation, followed by knowledge interpretation, dissemination, use, retention and refinement. The researchers have identified many key aspects of the KM system from process point of view as: acquire, collaborate, incorporate and experiment (Leonard, 1995); put emphasis on knowledge creating, transferring, storing, assimilating and taking advantage of that knowledge (Teece, 1998); focused on three aspects of knowledge that is to create, transfer and use (Spender, 1996); KM process is consisting on knowledge creation, storing, transferring and applying (Alavi and Leidner, 2001); KM is termed as a process for knowledge creation and manages the distribution and sharing of knowledge within and between each organization. Darroch’s definition of KM portrays that KM is made out of three main sections, which are knowledge acquisition, knowledge dissemination and knowledge responsiveness; As per (Nonaka and Takeuchi, 1995) KM consists of knowledge creation, knowledge storage, knowledge transferring and knowledge utilization.

On the basis of above literature KM philosophy grouped into knowledge creation, knowledge retention, knowledge transfer and knowledge utilization. These four elements in this study comprised the major concepts of KM.

However, knowledge creation is defined as the acquisition of new knowledge and updates the existing knowledge. The personal experience is required to share and disseminate the organizational knowledge. (Gold, Malhotra and Segars, 2001). When knowledge is created, the system of knowledge storage becomes important. The process of knowledge storing in a database affects the process of knowledge sharing and transferring (Lim and Klobas, 2000). Knowledge storage is defined as how an organization captures and stores the different units of knowledge in such a form that give various categories and labels, to the input data (Zack, 1999). Knowledge transfer is defined as the distribution of knowledge among all persons involved in the business processes (Lin and Lee, 2005). Knowledge utilization is defined as the business process that helps to access knowledge easily from the databases and apply it (Lin and Lee, 2005).
Reed, Lemark, and Mero (2000) provided a comprehensive analysis of various TQM models presented by known scholars from 1974 to 1996 and pointed out the common similarities on TQM practices. According to this review all well known quality experts are agreed on the importance of the following six key practices: cost reduction, leadership and top management commitment, training and education, teamwork and organizational culture and customer satisfaction. According to Dean and Bowen (1994), three TQM principles that are common in most quality frameworks are teamwork, continuous improvement and customer focus. Sila and Ebrahimpour (2002) on the basis of survey-based research on TQM in different countries and grouped the TQM practices under 25 categories. The first seven factors are considered to be the major elements of TQM these are customer focus and satisfaction, training and employee involvement, top management commitment, leadership and employee empowerment.

Furthermore, many studies have been conducted to measure the quality information and performance measurement. On the basis of literature it is found that no research clarifies the key elements of TQM that show the capabilities of TQM (Shenawy et al., 2007). Hence, due to inconsistency in the previous research it is difficult to identify the exact elements of TQM (Hoang et al., 2006). After complete assessment of TQM literature and comprehensive examination of past research, four constructs of TQM practices namely Leadership, Strategic planning, HR Focus and Customer focus are being used in this research study as shown in proposed framework (Figure 2).

3. RESEARCH FRAMEWORK

![Conceptual Framework](image)

Fig. 1: Conceptual Framework

The conceptual framework is developed to examine the relationship between KM philosophy, TQM practices and performance excellence as shown in Figure 1.

In this conceptual framework KM philosophy is considered as an independent variable and TQM practices are considered as mediating variable whereas performance excellence is considered as dependent variable. KM philosophy includes knowledge creation, knowledge retention, knowledge transfer and knowledge utilization whereas TQM practices includes leadership, strategic planning, HR focus and customer focus as shown in figure 2.

The present study is the only one that examines the relationship between KM philosophy, TQM practices and performance excellence.
4. INTEGRATION OF KM PHILOSOPHY IN TQM PRACTICES

4.1 KM Philosophy and Leadership:

Leadership can be defined as the process by which others are being influenced to achieve some kind of desired goals (De Jong and Den Hartog, 2007).

But for those firms which are based on Knowledge and Knowledge management, there is a need to bring a change in TQM elements particularly in leadership styles (Powell, 1995). Macneil (2001) put emphasis on the management leadership to develop the core competencies and skills at workplace by organizational learning, particularly by helping the workforce to develop knowledge management initiatives and employees are encouraged to apply their tacit knowledge to solve the problems.

Many studies have confirmed that senior managers are playing an important role in controlling the rate of success for KM behaviors (Wong, 2006; Horak, 2001; Holsapple and Joshi, 2000). Bryant (2003) mentioned that company’s mission, motivation systems and structures design for the various actions that supply the means to operate knowledge should come from management leadership. For the development and enhancement of shared learning ability in organizations the role of senior managers as a helper in supporting the practice of knowledge management in teams, namely knowledge creation, knowledge dissemination and knowledge sharing is very important (Ellinger and Bostrom, 1999).

Wong (2006) recommended that the management leadership should set an example to freely transfer their knowledge and establish a culture that encourages the creation and transferring of knowledge. In a study done by Storey and Barnett (2000), they found that the KM success is based on the efforts of management leadership for continuous support...
and practical application of Knowledge management behaviors such as knowledge utilization, knowledge creation, knowledge storage and knowledge transfer.

On the basis of above literature we can say that KM Philosophy is positively associated with leadership. Hence the following preposition is developed.

**P1:** Knowledge management philosophy has a direct positive effect on leadership.

### 4.2 KM Philosophy and Strategic Planning:

Strategic planning can be defined as the process managers use to formulate and implement strategies for providing the best value to the stakeholders and achieve the predetermined goals of the organization (Bounds et al., 1994). Carayannis et al. (2000) and Grant (1996) commented that a firm becomes competitive mainly due to four reasons; these are special knowledge of its human resource, the ability of a firm to create new knowledge and be innovative, and the strategic actions taken by the firm. Liebowitz (1999) mentioned that the strategy provides a foundation for how a firm can organize its capabilities and resources to attain its KM objectives and this strategy comes from Knowledge and knowledge management. Major source for the competitiveness of all firms is due to the efforts to associate KM philosophy to strategic planning (Chong et al., 2006). For pursuing KM initiatives towards organizational success the strategy along with vision and employees knowledge are working together (Wong, 2006). It was further investigated by Wong (2006) before making any significant investment to initiate a KM effort, management must develop the mission, vision and strategies. The American Productivity and Quality Centre (1999) made a study and concluded that those firms have more success which is developing their strategic plans in association with KM philosophy. Hence, based on this, for firms that wish to implement KM philosophy, it is essential to make sure that their knowledge programs are matching with the company’s missions.

On the basis of above literature we can say that KM philosophy is positively associated with Strategic planning. Hence, the following preposition is developed.

**P2:** Knowledge management philosophy has a direct positive effect on Strategic planning.

### 4.3 KM Philosophy and HR focus:

Human resource management (HRM) can be defined as the policies and practices one needs to carry out the people or human resource aspects of a management position including recruiting, screening, training, rewarding and appraisals (Dessler, 2000).

For successful knowledge management people need to be given sense of security, motivated through incentives, training and empowered with authority. Malhotra (2003, 2005) recommended that that humans and processes are essential for knowledge management. Humans compare, connect, conclude and derive knowledge from data to develop the processes. Similarly for the success of knowledge management it is necessary that all the interested parties of business should take part in decision making. People are regarded as the most important asset in the present knowledge based environment (Fang et al., 2005). Oltra (2005) mentioned that knowledge and human resources are considered the main force of a complicated business environment.
Alvesson (1993) recommended that the ultimate creator and owner of knowledge are the people. According to Lin (2007), it was found that those individuals who often enjoy to help others they at the same time improve their self knowledge. It means that knowledge is increased by sharing it with the support from senior management and proper organizational rewards, which could then lead to superior innovation. Briefly Lin (2007) concluded it as those firms which are encouraging the culture of sharing and collecting the knowledge by the mutual efforts of all the employees lead to increase in innovation performance. Many researchers put the emphasis on the significance of Human resource management as one way to increase knowledge transfer, in particularly in the form of technology expertise (Sparkes and Miyake, 2000), due to the reality that HRM and KM have become more broad and sophisticated (Gloet, 2006).

In accordance to Dougherty (2001), teamwork is helpful in sharing out the work, which as a result helps in the knowledge transfer within an organization. Numerous studies were done by different scholars and researchers to test the relationships between human resources and KM.

On the basis of above literature we can say that KM philosophy is positively associated with HR focus. Hence, the following preposition is developed.

**P3:** Knowledge management philosophy has a direct positive effect on HR focus.

### 4.4 KM Philosophy and Customer focus:

Customer focus is defined as the extent to which an organization continuously satisfies customer needs and expectations (Zhang, 2000) It is considered as one of the basic building blocks of TQM (Bank, 2000). Due to high competition in business environment, the major pressure on management of almost all of organizations is to focus on customer needs (Piercy, 1995). According to Hackman and Wageman (1995), one of the most frequently used TQM implementation practices is obtaining information about customers. Sila (2007) and Brah et al. (2000) both focused that the success of any organization in coming future is depending upon the satisfaction of its customers’ needs efficiently and effectively on continuous basis. The basic principle of TQM is to create a value for the customers (Mele and Colurcio, 2006; Woodruff, 1997). To exemplify the customer focus the knowledge is required for collecting the information about the expectations of the customers and then to distribute such information within the organization (Dean and Bowen, 1994). All the business processes that put emphasize on customer knowledge are part and parcel of intellectual assets of any organization (Bassi and Van Buren, 1999).

On the basis of above literature we can say that KM philosophy is positively associated with Customer focus. Hence, the following preposition is developed.

**P4:** Knowledge management philosophy has a direct positive effect on customer focus.

As mentioned in the literature above it is found that TQM practices in integration with KM philosophy are positively associated with performance excellence. Hence, the following preposition is developed.

**P5:** TQM practices in integration with KM philosophy have a direct positive effect on Organizational performance.
5. CONCLUSION

Many studies have been conducted to examine the relationship of TQM practices and KM behaviors with organizational performance on individual basis. This is the first study which investigates the combined effect of KM philosophy and TQM practices on performance excellence. The focus of this study is on those organizations which are large, complex and diversified. The proposed framework seeks to close the gap in literature for assessment of integration of KM philosophy in TQM practices for performance excellence. This is an initial study to examine the relationship of KM philosophy and TQM practices on performance excellence. As for improvements further empirical study should be conducted to prove the proposed prepositions significantly.

6. PRACTICAL IMPLICATION

This study is proposing that the top management can achieve the excellent performance if they used their resources for Knowledge management Philosophy along with TQM practices. By using this framework the practitioners can get a better understanding of how different aspects of knowledge management and total quality management in an organization contribute the excellent performance of an organization.

REFERENCES


GIS FRAMEWORK USING SDI FOR DISASTER MANAGEMENT

Umer Farooq*, Khalil Ahmed, Yousaf Saeed and Sagheer Abbas
School of Computer Sciences, National College of Business Administration and Economics, Lahore, Pakistan
*Corresponding Author: umer@ncbae.edu.pk

ABSTRACT

Spatial data and associated technologies have recognized to be important for effective collaborative decision making in disaster management. There are currently problems with ease of use, access and usage of reliable, up-to-date and accurate data for disaster management. It is very important aspect to disaster response as timely, up to date and precise spatial data important the current situation is paramount to successfully responding to in emergence. The up to date spatial data can provide the information about available resources, access to roads, damaged areas and required disaster response operations that should be accessible in short time. The collection of information and sharing among concerning departments is a challenge in disaster management. Therefor it is necessary to utilize appropriate framework and technologies to resolve current spatial data problems for disaster management. This paper objects to address the role of spatial data infrastructure as a framework for the real time monitoring of the earth and decision making for the disaster.

1. INTRODUCTION

More than 500,000 people were affected in various natural disasters all over the world in the last decade. Disaster affect our society in vary aspects. (Sutanta, Rajabifard and Bishop 2009). Gujarat India 2001, Kashmir Pakistan 2005 Java Indonesia 2006-7, Sichuan and Haiti China 2008-10, were the decade deadly earthquakes of the history. Thousands of people died, millions are injured and displaced. Huge financial and life lost were faced by the nations. Natural and man-mad disasters, such as earthquakes, floods, plane crashes, high-rise building, road accidents, terrorism incidents, fires or major nuclear failure are challenging jobs to public emergency services. Manage to such disaster in a fast and in time, the information regarding the location is vital necessity. Police, fire departments, public health, civil defense and other organization have to react not only efficiently and individually but also in a synchronized manner (Meissner, et al. 2002).

Generally, emergency happen variably and cause individual and organizations to shift their attenuations to deal with the circumstances instantly. Whenever disaster come to be large scale, all the limitation result a lack of integration and collaboration among all the involved organization being exposed and further compounds the negative consequence of the event. In the large-scale disaster the people who work together, they have not developed understanding of one another’s abilities; such coordination causes the wrong decision-making in the worst situation (M.Turoff 2008). The improvement in the remote
sensing technology can provide favorable condition for extract useful data for disaster prevention and control. There are four different types of disasters, natural disaster, environmental emergencies, complex emergencies and pandemic emergencies. Any disaster can affect essential services like, health care, electricity, transportation and communication. Disasters influence the people long, after the instant effect has been mitigated. The relief and recovery needs proper planning. Poor strategies have negative effects not only on the disaster victim’s but also on donors and relief agencies. Managing disaster properly require the no of activities, hazard prevention, preparedness, relief and recovery.

The spatial data and correlated technologies are effective collaboration for the decision-making in disaster management. The monitoring of the areal extant of earth is important for climatological changes and disaster management. The field of disaster management has significantly benefited from recent advancement and related technology. Especially GIS, remote sensing, and internet have had a significant influence, and presently are being used in verity of ways during all stages of disaster management. The association of these technologies to create the spatial decision support system offers the new dimensions for disaster management, particularly during the early response phase, when the change detection ability is most urgently required (S.Herold, 2005).

Both human and economic loss can be reduced by using the spatial data and associated technologies. This can be helpful to quick recover of the disaster and plan well the challenges of rapidly recover of disaster area.

2. LITERATURE REVIEW

GIS, remote sensing and related technologies widely used in monitoring, predicting, disasters early warning system, loss assessment, rescue activities, and aid crisis management for natural disaster management systems. The effectiveness and usefulness of the technology could be the prime factor if they were properly incorporate into disaster management system (Wang and Li 2008).

The literature shows the importance of the GIS based decision support systems are the growing in the field disaster management. The applications are focused on technical measurement of GIS software system, spatial data structure, capture techniques, hazards modeling and its spatial structure (Kakumoto, Higashihara and Sasaki 2008).

J. Goncalo et al present SAGE-B for the dams and the elements associated to an emergency in case of dam-break flood. SAGE-B support all related data to dams and the constituent supplementary to an emergency in case of dam-break flood. The information associated with the system that, population located in the area at risk, evocation rout plan and vehicles for the rescue.

Z. Andre and S. David proposed the real time decision support system for the risk and disaster management using GIS. The widely availability of spatial databases, low cost of software and the expertise of the risk managers in GIS are the reason to believe the utility of GIS systems for Natural hazard risk and disaster management systems. The researchers evaluates the application of GIS for cyclone disaster risk-management,
3. PROPOSED MODEL

The prevention and preparedness provides the protection from the massive destruction of disasters, especially form the natural disasters. The risk of life lost and injuries can be reduced by prevention measurements, early warning system, resources allocation, good evacuation plans and short time response to disaster. In this paper we suggesting emergency control system using GIS and spatial data integration. Data gathering, organizing and displaying logically is the key factor in all phases of the emergency control system to determine the scope of emergency. Emergency can affect all or number of government organizations. In the relief operation detailed information is required during the work. Every organization has data for their use like health care, police, fire brigade and other rescue organization. The using GIS in emergency control system emergency worker can share information in on place by maps and other data. The spatial data can be gathered, displayed and analyzed for the emergency operation. GIS allows the mechanism to centralize the spatial data and visually display maps, extract useful information during an emergency. The information need in disaster management can be divide in two categories, pre-order and post-order and emergency management activities can be grouped as planning, mitigation, preparedness and response. The centralization of information will be useful for all these actives. The figure 1 shows ECS (Emergency Control System) sharing the information with all organization. Major challenge associated with the geographical systems is the connection with all people, organizations, services, and easily accessible when needed. For example even the first response of the mobile user in crisis situation, the user has to be able to access up-to-date information from the disaster management service. Figure-2 explains the proposed solution for disaster management system. Real time analyzing of the earth and change detection in the stored data on the GIS content server is the key component for emergency control personals. The communication between the ECS and GIS content server through the data analyzer could be wireless or cable based data synchronization communications. Both methods facilitate two way communications with several middleware synchronization software tools. The all connected organization with ECS, can plan well for response early in disaster. Various methods can be used for real-time monitoring and change detection in GIS content server data. Robin et al. present unsupervised sub-pixel change detection method using image series. Most of the change detection methods are based either on classification process, joint classification or on fusion classification (Robin, Moisan and Hegrat 2009). Markov random fields method used for model based approaches (Melas and Wilson 2002), vector analysis threshold method for histogram, classification and segmentation. Real-time emergency management will need to access the information immediately for warning systems, evacuation, or responder. There for emergency system can tolerate minimum time to response the request made to the GIS system. The ECS can response as real time emergency tasks as pragmatic maps with GPS function for navigation and resources allocation according to the emergency situations.
Fig. 1: Emergency Control System Information Flow

Fig. 2: ECS Framework for Disaster Management
4. CONCLUSION

Spatial data and associated technologies have recognized an important for effective collaborative decision making in disaster management. The field of disaster management has significantly benefited from recent advancement in related technology. The purposed framework can be help full for real time emergency control system. The ECS can reduce the risk of life lost and injuries by early warning system, resources allocation, GPS based navigation for evacuation and quick response to disaster. The centralization of information provides the facility to all organizations to participate in emergency.

REFERENCES

RISK MANAGEMENT FOR CNG BASED VEHICLES

Yousaf Saeed, Khalil Ahmed, M. Saleem Khan and Umer Farooq
School of Computer Sciences, National College of Business Administration and Economics, Lahore, Pakistan
*Corresponding Author: usafonline.email@gmail.com

ABSTRACT

This research paper discusses the risk assessment of bi-fuel vehicles having Compressed Natural Gas (CNG) kits. The focus is on gas tanks of CNG kits where two aspects are considered: First, filling CNG tanks with low pressure not only prevents vehicles for covering longer distances but financially it is a loss as well. Secondly, filling CNG tank of a vehicle at a gas station maximizes the risk of exploding when it exceeds a certain threshold level. In this paper, a framework has been defined to explain these two aspects.

1. INTRODUCTION

With the intense usage of a number of natural resources in the world, shortage issue is going to rise in the near future. As a matter of fact, alternative measures are necessary to keep the world going and to make it green. Same is in the case with vehicles running on gasoline. Natural gas, in its natural form does not smell, it is odorized to a certain extent so that it can be detected during possible leakage. It consists of about 90 percent methane.

Today, Compressed Natural Gas (CNG) is used as an alternative fuel source in many countries around the world. On the other hand, it is more environmental clean product as defined in [1].

By 2010, there were 12.6 million vehicles worldwide. Pakistan being on top having 2.74 million vehicles, Iran being the second having 1.95 million, Argentina having 1.9 million, Brazil 1.6 million, India 1.1 million vehicles and so on as shown in figure 1 and is defined in [2].

Regarding CNG cylinders depletion, it gets warm when refilled. If the gas is filled with pressure, the molecules of the gas increases and rises the temperature inside; however, after a while the gas adopts the temperature of the environment. These properties also have effect on the storage capacity of the cylinder when it is refueled. Like if the temperature increases inside the cylinder, the pressure increases as well. That is why the dispensers at CNG stations automatically stop dispensing CNG when a pressure of 200 bar is reached. Theoretically, if 20 kg is the capacity of a certain cylinder under standard conditions like 200 bar at 20° C, practically the cylinder will not be filled with the exact amount of 20 kg but a bit less than 20 kg.
2. LITERATURE REVIEW

Regarding Compressed Natural Gas in vehicles, research has been carried out by considering different aspects of vehicles. Energy sources are focused primarily by the researchers as they are now used in each and every area.

Neural network and Fuzzy control system are widely used in a number of technological aspects and known. Zhang, Jiang, Xia and Zhou investigated Fuzzy Neural Hybrid Controller (FNHC). Gradient decent learning algorithm is used for the controller’s robustness and adaptability. Fuzzy Neural Hybrid Controller has been implemented on a six cylinder engine lean-burn CNG engine to regulate the air fuel ratio and defined in [3].

Robert Zalosh describes about the possible risks regarding a vehicle’s CNG fuel tank and prevention. Fuel tank design, fabrication and testing standards are focused. Different incidents due to gas tanks are reviewed along with different testing methods and defined in [4].

3. PROPOSED MODEL

The proposed architecture for CNG vehicles consist of several modules that include a Pressure Sensing Unit which senses the pressure of natural gas to find out how much it is compressed, this unit is intelligent enough to distinguish the gas on the basis of pressure and categorize it into four types i.e. low, normal, ideal and high. Low category is the low pressure of natural gas which is not suitable for CNG vehicles. Although the vehicle is able to travel when filled with low pressure but the life of the cylinder is compromised to certain extent. High category is the high pressure of natural gas. It is also not suitable to fill the gas cylinder with high pressure. It is more dangerous than cylinder with low pressure and many CNG cylinders are marked as bad due to high pressure.
Ideal category deals with the ideal pressure for vehicle’s CNG cylinder. It is the most appropriate pressure considered by various standardized organizations. It also varies from cylinder to cylinder in vehicles and categorized as Type-1, Type-2 and Type-3. Normal category too is considered as good pressure, however, there is a slight difference between normal and ideal pressures categories being ideal the most significant followed by the normal pressure.

The Pressure Management Unit includes the key Priority Unit that gives priority to the gas regarding its pressure. In this architecture, highest priority is given to the ideal pressure which ranges between 225 to 230 bars. The second priority is given to the normal pressure which ranges between 220 to 225 bars. Both of these ranges are defined in Table 2. Third priority varies and is dependent on the gas cylinder status of the vehicle. According to the company policy, vehicle’s gas cylinder has to be checked after every five years, however, inspection is carried out after every year. If the gas at the station is sensed as low or high in pressure, the driver is asked about the inspection status, if the inspection has been done and the cylinder is in good position, high pressure at that moment is provided otherwise low pressure is provided. These tasks are handled by the Cylinder Inspection Unit. Lastly, the Gas Flow Unit is used as a gas dispenser into the vehicle’s cylinder.
Pressure of the gas can be find out with the help of the following formula,

\[ \text{Pressure} = \frac{\text{Force}}{\text{Area}} \]

Also, the initial pressure, initial volume, final pressure and final volume can be find using Boyle’s law. According to Boyle’s law, the pressure and volume is inversely proportional in the temperature and amount of gas is kept constant.

\[ P_1 V_1 = P_2 V_2 \]

Different levels of fuel in kilogram are defined into a number of categories. The levels start from 1 to 8 starting from 0 to 480 respectively as shown in Table 1. The reason for extending the level up to 8 is due to the reason that the gas cylinder is filled double its capacity by the standard checking authority over every five years. During this process if the cylinder expands at a certain position, it is discarded and if not, it is considered safe for usage.

<table>
<thead>
<tr>
<th>Fuel in k.g.</th>
<th>Level 1</th>
<th>0 – 60</th>
</tr>
</thead>
<tbody>
<tr>
<td>Level 2</td>
<td>60 – 120</td>
<td></td>
</tr>
<tr>
<td>Level 3</td>
<td>120 – 180</td>
<td></td>
</tr>
<tr>
<td>Level 4</td>
<td>180 – 240</td>
<td></td>
</tr>
<tr>
<td>Level 5</td>
<td>240 – 300</td>
<td></td>
</tr>
<tr>
<td>Level 6</td>
<td>300 – 360</td>
<td></td>
</tr>
<tr>
<td>Level 7</td>
<td>360 – 420</td>
<td></td>
</tr>
<tr>
<td>Level 8</td>
<td>420 - 480</td>
<td></td>
</tr>
</tbody>
</table>

Pressure can be measured either in bar or Psi, however, pressures have been grouped together starting from 205 bar being low and raised gradually to a high pressure of 245 bar as shown in Table 2.

<table>
<thead>
<tr>
<th>Pressure in bar</th>
<th>Very Low: 205 - 210</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low</td>
<td>210 – 215</td>
</tr>
<tr>
<td>Medium Low</td>
<td>215 – 220</td>
</tr>
<tr>
<td>Normal</td>
<td>220 – 225</td>
</tr>
<tr>
<td>Ideal</td>
<td>225 – 230</td>
</tr>
<tr>
<td>Medium High</td>
<td>230 – 235</td>
</tr>
<tr>
<td>High</td>
<td>235 – 240</td>
</tr>
<tr>
<td>Very High</td>
<td>240 - 245</td>
</tr>
</tbody>
</table>

To represent my work with the proposed architecture, MATLAB has been used for simulation. We have taken one input which is fuel with four outputs which are ideal, normal, low and high and shown in figure 3.
Eight membership functions have been taken for fuel as represented by different levels starting from 1 to 8 with a range starting from 0 to 480 in kg and shown in figure 4.

Again eight membership functions have been taken for pressure. The range starts from 180 to 260 bar, 180 being very low and 260 being very high which is shown in figure 5.

Different rules have been defined on the basis of membership functions considering both the input and outputs while setting the fuel ratio to 226 bars as shown in figure 6.
The output results are generated on the basis of member functions and its defined rules. The straight line in figure 7 indicates the ideal pressure before the gas is going to be filled in the vehicle CNG cylinder. Normal pressure is indicated in figure 8 which is the second best option after ideal pressure. Low and high pressures are indicated in figure 9 and 10 respectively. Both are dependent on the cylinder inspection status unit.
4. CONCLUSION

CNG vehicles are used as an alternative to other fuel sources mainly gasoline. It is a good energy source for vehicles that generates less pollution, lower maintenance cost, economically suitable and is free from spill. With all these advantages, the current energy source has a downside in the form of safety risk. The proposed architecture will provide safety not only to the passengers but to the vehicle itself, gas station, people around and to the environment.

REFERENCES


CIRCULATING SERUM CONCENTRATIONS OF ADIPOCYTOKINES IN OBESE AND TYPE 1 DIABETIC CHILDREN AND YOUNG ADOLESCENTS

Anila Jaleel*, Bilal Aheed, Faisal Yaseen, Sana Yousuf, Harris Hashim and MS Farah
Department of Biochemistry, Ziauddin University, Karachi, Pakistan
*Corresponding Author: aneelajaleel@hotmail.com

ABSTRACT

Background:
Adipocytokines like Leptin, adiponectin, visfain and resistin play important role in the regulation of intermediate metabolisms.

Objective of the study:
The aim of this study was to evaluate the circulating levels of adiponectin and resistin in obese and type 1 diabetic children and young adolescents

Methods:
A total of 89 subjects between the age of 05-18 years were selected and divided into three groups of 30 controls 29 obese and 30 type I diabetic subjects. Serum resistin, insulin and adiponectin were measured using ELISA and random blood sugar by glucose oxidase (Standard Kit method).

Results:
Out of 90 study subjects, 30 were normal healthy subjects, 29 obese and 30 type 1 diabetic patients. Serum insulin level in controls was 7.28 ± 0.96 which increased significantly in obese individuals (P<0.01) 15.44 ±3.92 and does not change significantly in type I diabetic subjects compared with controls (6.35 ± 1.17). Mean HOMA-IR was 1.69 + 0.45 for control group but significantly increased P < 0.05) in obese children (3.99+ 0.94).However it was 1.79 ± 0.18 in type 1 diabetic patients which remains nonsignificant compared with controls.. Mean serum adiponectin decreases significantly (P, 0.05) in obese (5.96 ± 1.02) compared with controls (9.98 ± 0.79). However it increases in type I diabetic subjects but it was non significant. Serum resistin increases significantly P< 0.05 (5.87 ± 3.36) in obese individuals and in type 1 diabetic subjects (12.38+6.24) compared with controls (4.91+ 2.85). Inverse significant correlation was found between adiponectin and resistin in obese subjects. However positive correlation was found between adiponectin and resistin in type 1 diabetic patients.

Conclusion:
High HOMA score in obese children shows that they are insulin resistant and will develop type 2 diabetes in future. This is also supported by low serum level of adiponectin and high resistin level in obese children showing the loss of anti-inflammatory and increased pro-inflammatory role of adipocytokines in these subjects. However increased resistin levels in type 1 diabetic patients shows increased
proinflammatory effect. Moreover increased adiponectin concentrations in type I diabetic subjects needs to be further explored in large scale studies.

KEYWORDS
HOMA-IR, Adiponectin, Resistin, Insulin.

INTRODUCTION

Adipocytes synthesize and secrete several cytokines including tumor necrosis factor \(\alpha\), resistin, visfatin, leptin and adiponectin, the latter being one of the most abundant adipose tissue–specific adipocytokines. In the last years, an attractive hypothesis has emerged proposing that those cytokines produced by adipose tissue may be responsible for insulin resistance in obesity\(^1\,^2\).

Obesity is an emerging global public health challenge in developing as well as developed countries. Several environmental factors contribute to the obesity epidemic that is now being observed among children: the sustained excess of energy-dense foods with refined carbohydrates and high saturated fat (the age groups between 0 and 19 years are those with the highest intake of saturated fats per total calories consumed), as well as insufficient consumption of fruit and vegetables. The impact of these nutritional factors is further aggravated by an increasingly sedentary lifestyle—attributed in part to urbanization, which limits the opportunities for physical activity, and the predominance of electronic entertainment over physical activity. The rapidly rising burden of obesity with persistent levels of undernutrition among Indo-Asian children is a unique and complex challenge and represents a major threat to the healthcare services. Thus, there is a clear need to focus health policies on combating this rising epidemic of energy imbalance, which is shifting the pendulum towards overweight and obese children in Indo-Asian countries, while paying attention to the needs of the ones who are still undernourished. The prevalence of childhood and adolescent obesity has increased over the past three decades such that obesity is now a worldwide pediatric health risk factor.\(^3\) In parallel, the incidence of obesity-associated diseases such as Type 2 diabetes mellitus and dyslipidemia is rising dramatically. Because children are relatively free from comorbidity compared with adults, we examined the role of serum resistin and adiponectin levels as a marker of juvenile obesity and insulin resistance and also role in Type 1 diabetic children.

Present study was aimed to investigate the markers of inflammation in obese children as well as children with type 1 diabetes.

PATIENTS AND METHODS

Eighty-nine male subjects between 5 to 19 years of age from same socioeconomic status, were be recruited from the community. They were divided into three groups of 30 subjects each i.e. non obese non diabetic controls and diabetic type 1 while 29 obese subjects. Informed consent was taken from children or their parents and the ethical committee Ziauddin University approved the project.
Body weight, height, Body mass index (BMI) and waist hip circumference were measured by standardized procedures. Although body mass index (BMI) >25 indicates overweight and >30 defines obesity in adults, the diagnosis of overweight in children relies on age-adjusted percentiles. Obesity in the same time period was 1.6-fold. For the purpose of their study published in the present issue of Circulation, Woo and colleagues used the age-adjusted BMI cutoff points and defined overweight as a BMI between 21. and 23 and obesity as a BMI >23 in children⁴.

The estimate of insulin resistance by HOMA score was calculated with the following formula: fasting serum insulin (µU/ml) x fasting plasma glucose (mmol/l)/22.5. As described by Hedblad et al.⁵, patients with HOMA score values exceeding the 75th percentile (i.e., 2.0) were considered to have insulin resistance.

Excluded were individuals having diseases other than the diabetic complications (endocrinological disorders, inflammatory conditions like Rheumatoid arthritis, malignancy etc) were excluded from the study.

PROTOCOL

After assessing baseline parameters, the subjects from all the three groups were assessed for inflammatory cytokines, such as adiponectin, resistin and insulin. Serum samples for cytokines level were stored at –80°C until assayed. Serum concentrations of adiponectin, resistin and insulin were determined by quantitative enzyme assays. Fasting plasma glucose was determined by glucose oxidase method.

STATISTICAL ANALYSIS

Mean and standard error of mean was determined. The groups were compared using students t test in SPSS version 17. P value <0.05 was considered significant.

RESULTS

The study included 89 subjects divided in three groups of thirty each in healthy controls and diabetic type 1 children while 29 obese subjects. All three groups were matched for age and sex.

Table 1 shows BMI and waist hip ratio of the three groups. Waist hip ratio and BMI of the obese children were significantly increased compared with the controls and diabetic type 1 children.
Table 1:
Comparison of Waist Hip Ratio and BMI amongst Controls, Obese & Diabetic with Type I Children

The values are expressed as mean and standard deviation (S.D). Number of cases is given in parenthesis

<table>
<thead>
<tr>
<th></th>
<th>Healthy Controls (30)</th>
<th>Obese Subjects (29)</th>
<th>Type 1 Diabetic patients (30)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Waist hip ratio</td>
<td>0.87 ± 0.05</td>
<td>0.90* ± 0.05</td>
<td>0.84 ± 0.05</td>
</tr>
<tr>
<td>BMI</td>
<td>16.25 ± 1.17</td>
<td>24.37*$ ± 7.26</td>
<td>17.09 ± 1.77</td>
</tr>
</tbody>
</table>

* P< 0.05 Significant compared with controls
* $ P< 0.05 Significant compared with controls and diabetic patients

Table 2 shows the plasma concentrations of insulin, adiponectin and resistin in three groups. Insulin resistance calculated by HOMA Score is also shown. Insulin levels and insulin resistance increases significantly in obese children compared with the other two groups. Adiponectin level decreases significantly in obese children compared with the other two groups. However adiponectin level increases in Type 1 diabetic patients. Resistin level increases significantly in obese and type 1 diabetic children compared with controls.

Table 2:
Comparison of Serum Glucose, Insulin, Adiponectin and Resistin amongst Controls, Obese and Diabetic Type I Children

The values are expressed as mean and standard deviation (S.D). Number of cases is given in parenthesis

<table>
<thead>
<tr>
<th></th>
<th>Healthy Controls (30)</th>
<th>Obese Subjects (29)</th>
<th>Type 1 Diabetic patients (30)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Glucose(mmo/l)</td>
<td>5.36 ± 0.92</td>
<td>5.95 ± 0.97</td>
<td>6.77 ± 0.51</td>
</tr>
<tr>
<td>Insulin(uIU/ml)</td>
<td>7.27 ± 0.96</td>
<td>15.58* ± 3.06</td>
<td>6.34 ± 1.17</td>
</tr>
<tr>
<td>HOMA IR</td>
<td>1.74 ± 0.45</td>
<td>4.04* ± 0.90</td>
<td>1.79 ± 0.17</td>
</tr>
<tr>
<td>Adiponectin (ng/ml)</td>
<td>9.97 ± 0.78</td>
<td>6.02* ± 1.02</td>
<td>10.83*$ ± 1.73</td>
</tr>
<tr>
<td>Resistin (ng/ml)</td>
<td>4.91 ± 2.85</td>
<td>5.74* ± 3.34</td>
<td>12.38*$ ± 6.23</td>
</tr>
</tbody>
</table>

* Significant compared with controls
$ Significant compared with Obese subjects
Significant inverse correlation was found between resistin and adiponectin ($P = 0.000$) in obese subjects. However significant positive correlation was found between adiponectin and resistin in type I diabetic patients subjects ($r = 0.367 \ P = 0.000$).

DISCUSSION

Our results demonstrate significantly increased concentrations of resistin and decreased levels of adiponectin in obese subjects. However resistin and adiponectin increases significantly in diabetic Type 1 patients. Serum glucose and insulin levels are increased significantly in obese subjects resulting in insulin resistance in this group calculated by HOMA score. Adiponectin is an anti-inflammatory cytokine which has protective role. Its level decreases as a result of inflammation and decreased adiponectin concentration is good candidate cause of endotheliopathy in insulin resistance and Type 2 diabetes. Obese children are shown to exhibit insulin resistance demonstrating underlying inflammation and risk factor for diabetes and ischemic heart disease, resulting in decreased adiponectin levels. Several studies have demonstrated decrease in protective role of adiponectin in obesity and insulin resistance and diabetes type 2 possibly due to neurohormonal activation and increased oxidative stress and infiltration of myocytes with free fatty acids. Diabetic type 1 patients showed increased concentration of adiponectin, which have been associated with increased severity of disease and adverse outcomes in patients.

Resistin may be an important link between increased fat mass and insulin resistance. In mice, resistin is expressed predominantly in white adipose tissue and is detectable in serum, suggesting that it is secreted by adipocytes and acts at distant sites. The findings in obese mice that serum levels of resistin are markedly increased and are decreased by rosiglitazone and other thiazolidinediones that increase sensitivity to insulin point to resistin as a mediator of insulin resistance. In an adipocyte cell line, resistin inhibits insulin-stimulated glucose uptake and antibodies against resistin enhance glucose transport in these cells, suggesting that endogenous resistin has autocrine effects.

CONCLUSION

Resistin and adiponectin appears to play an important role in obesity and insulin resistance in children and adolescents. The role of these cytokines in type 1 diabetics needs to be further studied.

ACKNOWLEDGEMENT

We owe thanks to Higher Education Commission project No 1357 for financial assistance.

REFERENCES


OPERATING SYSTEMS SCHEDULING A NEW PERSPECTIVE

Irshad Ullah
Department of Computer Science, Lahore University of Management Science, Lahore, Pakistan
Email: irshadullah79@gmail.com

ABSTRACT

Scheduling the resources within the operating systems is a critical and complicated issue. In this research an attempt has been made to give a new direction to the scheduling. Its may be that the scheduling is based in such a way that on the arrival of first process the resources have been allocated to that process and make it run on the basis of first comes first serve basis. While the coming process to be keeping in a Queue. And now to make such an arrangement to check that either the process is CPU bounded or I/O bounded. After that the scheduling has to be done on the quantum expiration or preemption and shortest job first basis.

KEY WORDS

Operating Systems, Resources, Scheduling, Hybrid, Process

1. INTRODUCTION

Operating systems are the software that makes the hardware usable. Hardware provides “raw computing power.” Operating system makes the computing power conveniently available to users, by managing the hardware carefully to achieve good performance [31].

Operating systems can also be considered to be managers of the resources. An operating system determines which computer resources will be utilized for solving which problem and the order in which they will be used. In general, an operating system has three principal types of functions. (i) Allocation and assignment of system resources such as input/output devices, software, central processing unit, etc. (ii) Scheduling: This function coordinates resources and jobs and follows certain given priority. (iii) Monitoring: This function monitors and keeps track of the activities in the computer system. It maintains logs of job operation, notifies end-users or computer operators of any abnormal terminations or error conditions. This function also contains security monitoring features such as any authorized attempt to access the system as well as ensures that all the security safeguards are in place [11].

2. RELATED WORK

In the 1940s, the earliest electronic digital systems had no operating systems. Computers of this time were so primitive compared to those of today that programs were often entered into the computer one bit at a time on rows of mechanical switches.
Eventually, machine languages (consisting of strings of the binary digits 0 and 1) were introduced that sped up the programming process [24]. The systems of the 1950s generally ran only one job at a time. It allowed only a single person at a time to use the machine. All of the machine’s resources were at the user’s disposal. Users soon realized that they could cut down the amount of time wasted between the jobs, if they could automate the job-to-job transition. First major such system, considered by many to be the first operating system, was designed by the General Motors Research Laboratories for their IBM 701 mainframe beginning in early 1956 [7]. Its success helped establish batch computing 1960s [14]. The systems of the 1960s were also batch processing systems but they were able to take better advantage of the computer resources by running several jobs at once. The operating systems of the 1960s, while being capable of doing multiprogramming, were limited by memory capacity. This led to the various designs of multiprogramming such as variable position multiprogramming that helped to utilize the storage capacity much more efficiently [22]. MIT’s Department of Electrical Engineering was one of the pioneers of the timesharing system under the guidance of John McCarthy, Robert Fano and Fernando Corbato.

Since 1957, it had been running a computer IBM 704 in a batch-processing mode. However, the instructions of programming and the development of software were very difficult given the long turnaround time, the time between the submission of a job and the return of results, of hours and even days. This motivated them to develop a 8 system that would reduce the turnaround time substantially. This led MIT to implement the first timesharing system in November 1961 In April 1964, IBM introduced its new generation of mainframe computers, System/360. It was so named because it was aimed at full circle of customers, from business to science [16], [31]. At the peak, more than 1,000 people at Poughkeepsie were working on OS/360. These included programmers, technical writers, analysts, secretaries and assistants and all together some 5,000 staff-years went into design, construction and documentation of OS/360 between 1963 and 1966 [16].

People turnover often resulted in large numbers of software modules being scrapped and then rewritten by new people because the existing modules could not be understood (Brooks, 1975). The development model viewed the software writing process not as a once-and-for-all construction project, like the way IBM approached OS/360 project, but as a more organic process, like the building of a city. Thus, software would be conceived, specified, developed and implemented – then it can be improved over from time to time, with added “bells and whistles” [28].

The 1970s saw several significant development The TCP/IP (Transmission Control Protocol/Internet Protocol) started to become widely used Communications in local area networks were made practical and economical by the Ethernet standard developed at Xerox’s Palo Alto Research Center [31].

Other aspects of computer and network security such as viruses and hacking had increasingly challenged the operating systems. As a result, the design of a secure operating system received top priority at that time [31]. In 1975, two of these computer buffs Bill Gates and his childhood friend Paul Allen decided to write BASIC, in order to take advantage of these opportunities. They decided to develop BASIC language for Altair [25].
In 1977, a manufacturer of Altair-clone, IMSAL approached Gary Kildall to use CP/M for its products [25]. The watershed event in this decade starts with IBM’s entry into the personal computer market. As soon as the decision had been made, IBM moved with remarkable speed. [2].

IBM did legitimate the personal computer [9] during 1982-1983, the IBM personal computer became an industry standard. One company that benefited the most out of this was Microsoft. Almost every model of IBM PC and its clones were supplied with its MS-DOS operating system. As hundreds of thousands and eventually millions of machines were sold, money poured into Microsoft. By the end of 1993, half a million copies of MS-DOS had been sold, netting $10 million [31] This revenue stream allowed Microsoft to diversify into computer application software without having to rely on external venture capital. It also allowed Microsoft to cross-subsidize some of the software that initially did not succeed. For example, in mid-1983, Microsoft began to develop a word processing software package called Word. That product was released in November 1983 with a publicity splash which included distribution of some 450,000 diskettes demonstrating the program in the PC World magazine. Even so, Word was initially not a successful product and had a negligible impact on the market leader at that time, WordStar. But, the cash flow from the MS-DOS allowed Microsoft to continue to market the product at a loss until the opportunity came later to bundle it properly with its new generation of operating systems, Windows [5].

One major deficiency of MS-DOS was that it was not very easy to use Microsoft started its own graphical user interface (GUI) project in September 1981 Despite its successes and the dominating position of the Windows operating system, Microsoft faces several challenges. One is the US Department of Justice’s lawsuit against Microsoft charging that it had used its dominating position illegally. While it had lost in District Court, the case is currently pending in the Appeals Court. The other challenges have to do with the future of the operating system as such. The advent of the Internet has opened up new possibilities and challenges. First, there are open-source systems like Linux, which is available freely online to anybody who wants to view, download or adapt it. This clearly threatens Microsoft’s dominating position. Second, the Internet may provide a platform in which operating systems may become much less important. In this rapidly changing environment of computing and information technology, it is extremely difficult to say which direction the operating systems will Take.

3. PROPOSED APPROACH

In this paper we consider one function of the operating systems that is called Scheduling. To schedule the work of the CPU is a very challenging Job. As here are different methods for the scheduling of the CPU. Every method has its advantages and shortcomings. So here in this paper we present a new method.
By this when a process is arrived in the CPU the CPU will issue the resources on first come first serve basis. While on the arrival of the new process if that process is finish it will allocate the resources on first come first serve basis. Otherwise it will create a Queue for the coming processes and change the policy to timesharing or preempt the process to make the execution of the remaining process possible. It has also to see that’s either the processes or CPU bound or I/O bound. Mean while by preempting the process and taking the resources it has to consider the shortest job first policy.

4. CONCLUSION AND FUTURE WORK

In this research a new method for the scheduling policy is proposed. In future work can be done to make it more clear and understandable. Also some implementation has to be made to check its correctness and reliability.

5. REFERENCES


   www.computinghistorymuseum.org/teaching/papers/research/history_of_operating_system_moumina.pdf
SUDOKU SQUARE DESIGN, LAYOUT AND ANALYSIS

Muhammad Aslam\(^1\) and Munir Ahmad\(^2\)

\(^1\) Department of Statistics, Forman Christian College
(A chartered University) Lahore, Pakistan.
Email: profmaslam@hotmail.com

\(^2\) National College of Business Administration & Economics, Lahore, Pakistan. Email: drmunir@ncbae.edu.pk

ABSTRACT

Sudoku square design is a special type of Latin square design. In addition to rows, columns and treatments another source of variation called blocks is introduced. The name Sudoku is taken from Sudoku puzzle, a popular number placement puzzle. An \((m_1 \times m_2)^2\) grid is partitioned into \((m_1 \times m_2)\) spatially compact squares or rectangles (blocks), each containing \((m_1 \times m_2)\) experimental units and \((m_1 \times m_2)\) treatments are applied to these experimental units such that each treatment appears once in each row, once in each column and once in each block.

In this paper the layout, randomization and analysis of Sudoku Square Design is discussed in the light of the fact that the blocks are not orthogonal to rows and columns. A linear statistical model is proposed on the grounds that the blocks are the joint effect of the vertical grids and the horizontal grids.

KEYWORDS

Magic Latin Square; Grechte; Sudoku; horizontal grids; vertical grids; orthogonal.

1. INTRODUCTION

A 4 x 4 Latin square with balanced corners may be the first reference to Sudoku designs\(^7\). There are examples of \((r \times s)^2\) Latin Square, divided into \(k = r \times s\) squares or rectangles, called magic Latin square\(^4\). We have a special type of Latin square, called “gerechte” (a German word meaning fair or just)\(^3\). An \((m_1 \times m_2)^2\) grid is partitioned into \((m_1 \times m_2)\) spatially compact regions, each containing \((m_1 \times m_2)\) experimental units and \((m_1 \times m_2)\) treatments are applied to these experimental units such that each treatment appears once in each row, once in each column and once in each region. Bailey et al. (1990) have dealt with statistical analysis of gerechte (Sudoku) designs\(^2\). Aslam (2008) in his M. Phil dissertation renamed this design as Sudoku Square Design after Sudoku puzzle\(^1\). Analyzing Sudoku Design like Latin Square is not justified because blocks are not orthogonal to rows and columns\(^5\).
THE DESIGN STRUCTURE AND LAYOUT

Let us have a \((m_1 \times m_2)^2\) Sudoku Square Design; having \((m_1 \times m_2)\) rows, \((m_1 \times m_2)\) columns and spatially compact blocks, where \(m_1\) and \(m_2\) are integers greater than 1. We restrict our discussion to the cases where the blocks are rectangles or squares. We give below the structure of \((m_1,m_2)^2\) Sudoku Square Design for \(m_1 = m_2 = 2\), for \(m_1 = 2\), \(m_2 = 3\), for \(m_1 = 3\), \(m_2 = 2\) and for \(m_1 = m_2 = 3\).

<table>
<thead>
<tr>
<th>Horizontal Grid (1)</th>
<th>Vertical Grid.(1)</th>
<th>Vertical Grid. (2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>row 1</td>
<td>A</td>
<td>B</td>
</tr>
<tr>
<td>row 2</td>
<td>D</td>
<td>C</td>
</tr>
<tr>
<td>row 3</td>
<td>B</td>
<td>A</td>
</tr>
<tr>
<td>row 4</td>
<td>C</td>
<td>D</td>
</tr>
</tbody>
</table>

A \((2 \times 2)^2\) Sudoku Square Design

<table>
<thead>
<tr>
<th>Horizontal Grids</th>
<th>Row\Col.</th>
<th>Vertical Grids</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3</td>
</tr>
<tr>
<td></td>
<td></td>
<td>4</td>
</tr>
<tr>
<td></td>
<td></td>
<td>5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>6</td>
</tr>
</tbody>
</table>

A \((2 \times 3)^2\) Sudoku Square Design
Layout of a Sudoku Square Design for $m_1 = 3$, $m_2 = 2$, $n = (m_1 \times m_2) = 6$

<table>
<thead>
<tr>
<th>Horizontal Grids</th>
<th>Row\Col.</th>
<th>Vertical Grids</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>A</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>B</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>C</td>
</tr>
<tr>
<td>2</td>
<td>4</td>
<td>D</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>E</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>F</td>
</tr>
</tbody>
</table>

A $(3 \times 2)^2$ Sudoku Square Design

Layout of a Sudoku Square Design for $m_1 = 3$, $m_2 = 3$, $n = (m_1 \times m_2) = 9$

<table>
<thead>
<tr>
<th>Horizontal Grids</th>
<th>Row\Col.</th>
<th>Vertical Grids</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>A</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>D</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>G</td>
</tr>
<tr>
<td>2</td>
<td>4</td>
<td>B</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>H</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>E</td>
</tr>
<tr>
<td>3</td>
<td>7</td>
<td>C</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>I</td>
</tr>
<tr>
<td></td>
<td>9</td>
<td>F</td>
</tr>
</tbody>
</table>

A $(3 \times 3)^2$ Sudoku Square Design
RANDOMIZATION

The basic principle of randomization applies to Sudoku Square Design as to any other design.

It necessitates the application of the treatments to the experimental units at random subject to the restriction that a treatment can only occur once in each row, column and block. The basic principle is that “each plot has an equal probability of receiving any of the possible treatments, and each pair of plots not in the same row, column or block has the same probability of being treated alike.” The computer softwares are available, which can be used to select one solution to the Sudoku puzzle at random and then permuting all rows, all columns, all blocks and all treatments with the restriction that each treatment must occur only once in each row, column and block. The treatments may be allotted letters A, B, C, etc. or numbers 1, 2, 3, etc. entirely at random to further ensure randomization. If the experimental area is actually a square shaped piece of land divided into \( n^2 \), \((n = m_1 \times m_2)\) square shaped plots, it is possible to transpose rows and columns, provided the blocks remain unchanged by this action.

STATISTICAL ANALYSIS

Sudoku Square Design cannot be analyzed like the Latin Square design (See Mo et al) because the blocks are not orthogonal to rows and columns and some contrasts on blocks are also contrasts on rows and columns.

In fact the \( i^{th} \) row effect is made up of two components namely the \( g^{th} \) Horizontal Grid effect and the effect of \( i^{th} \) row effect within \( g^{th} \) Horizontal Grid. Similarly the \( j^{th} \) column effect is made up of two components namely the \( h^{th} \) Vertical grid effect and the effect of \( j^{th} \) column within \( h^{th} \) Vertical Grid. The blocks are created at the cross of the Horizontal Grids and the Vertical Grids. Hence the block effect is in fact joint effect (interaction) of Horizontal Grids and Vertical Grids.

Hence we propose the following fixed effect model:

\[
Y_{ijk(gh)} = \mu + \tau_k + H_g + R_{i(g)} + V_h + C_{j(h)} + B_{gh} + \varepsilon_{ijk(gh)}
\]

where \( g=1,2,\ldots,m_2, \ i=1,2,\ldots,n, \ h=1,2,\ldots,m_1, \ j=1,2,\ldots,n, \ k=1,2,\ldots,n, \ n= m_1 m_2 \)

\( Y_{ijk(gh)} \) = an observation belonging to \( i^{th} \) row, \( j^{th} \) column \( k^{th} \) treatment, \( g^{th} \) Horizontal Grid, \( h^{th} \) Vertical Grid and \( (gh)^{th} \) block.

\( \mu \) = over all mean, \( \tau_k \) = effect of \( k^{th} \) treatment,

\( H_g \) = effect of \( g^{th} \) Horizontal Grid, \( R_{i(g)} \) = effect of \( i^{th} \) row in the \( g^{th} \) Horizontal Grid

\( V_h \) = effect of \( h^{th} \) Vertical Grid, \( C_{j(h)} \) = effect of \( j^{th} \) column in the \( h^{th} \) Vertical Grid

\( B_{gh} \) = effect of \((gh)^{th} \) block created at the cross of \( g^{th} \) Horizontal Grid and \( h^{th} \) Vertical Grid, \( \varepsilon_{ijk(gh)} \) = error or residual.
We assume that the effects are linear and additive and \( e_{ijk(gh)} \) are independently and identically distributed as \( N(0, \sigma^2) \) The constraints are: \( \sum \tau = 0 \), \( \sum H = 0 \), \( \sum R_{i(g)} = 0 \), \( \sum \nu_h = 0 \), \( \sum C_{j(h)} = 0 \), \( \sum B_{gh} = \sum (HV)_{gh} = \sum (HV)_{gh} = 0 \) We have:

\[
\sum_{j=1}^{n} y_{ijk(gh)} = R_i = \text{i}^{th} \text{ row total, } \sum_{i=1}^{n} y_{ijk(gh)} = C_j = \text{j}^{th} \text{ column total }
\]

\[
\sum_{i=(g-1)m_{1}+1}^{gm_{1}} \sum_{j=(h-1)m_{2}+1}^{hm_{2}} y_{ij} = T_{gh} = \text{(gh)}^{th} \text{ block total, } \sum_{i=1}^{n} y_{ijk(gh)} = T_k = k^{th} \text{ treatment total }
\]

\[
T_{g_{h}} = \sum_{h=1}^{m_{1}} T_{gh} = g^{th} \text{ horizontal grid total } T_{h} = \sum_{g=1}^{m_{2}} T_{gh} = h^{th} \text{ vertical grid total. }
\]

Then

\[
\bar{Y}_{i(\cdot)} = \frac{R_i}{n} = \text{i}^{th} \text{ row mean (or simply } \bar{Y}_{i} ), \bar{Y}_{j(\cdot)} = \frac{C_j}{n} = \text{j}^{th} \text{ column mean (or simply } \bar{Y}_{j} )
\]

\[
\bar{Y}_{k(\cdot)} = \frac{T_k}{n} = \text{k}^{th} \text{ treatment mean (or simply } \bar{Y}_{k} ), \bar{Y}_{g_{h}} = \frac{T_{gh}}{nm_{1}} = \text{g}^{th} \text{ Horizontal Grid Mean}
\]

\[
\bar{Y}_{h} = \frac{T_{h}}{nm_{2}} = \text{h}^{th} \text{ Vertical Grid Mean } \bar{Y}_{gh} = \frac{T_{gh}}{n} = \text{(gh)}^{th} \text{ Block Mean}
\]

\[
\sum R_i = \sum C_j = \sum B_{k} = \sum T_{k} = G = \bar{Y} = \frac{G}{n^2} = \text{the overall grand mean, } C.F. = \frac{G^2}{n^2} = \text{the correction factor.}
\]

The breakup of the total sum of squares [6]:

\[
\text{Total } SS = \sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{k=1}^{m_{1}} \sum_{g=1}^{m_{2}} (y_{ijk(gh)} - \bar{Y})^2
\]

\[
= \sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{k=1}^{m_{1}} \sum_{g=1}^{m_{2}} [(\bar{Y} - \bar{Y}_{g})^2 + (\bar{Y} - \bar{Y}_{h})^2 + (\bar{Y} - \bar{Y}_{j})^2 + (\bar{Y} - \bar{Y}_{k})^2 + (\bar{Y} - \bar{Y}_{gh})^2 - 2(\bar{Y} - \bar{Y}_{g})(\bar{Y} - \bar{Y}_{h})]
\]

\[
= n \sum_{g=1}^{m_{1}} \sum_{i=(g-1)m_{1}+1}^{gm_{1}} \sum_{h=1}^{nm_{2}} (\bar{Y} - \bar{Y}_{g})^2 + n \sum_{g=1}^{m_{1}} \sum_{j=1}^{m_{2}} (\bar{Y} - \bar{Y}_{j})^2 + n \sum_{g=1}^{m_{1}} \sum_{h=1}^{m_{2}} (\bar{Y} - \bar{Y}_{h})^2 + 2n \sum_{g=1}^{m_{1}} \sum_{h=1}^{m_{2}} (\bar{Y}_{gh} - \bar{Y}_{g})(\bar{Y}_{gh} - \bar{Y}_{h}) + \sum_{i=1}^{n} \sum_{j=1}^{n} (\bar{Y}_{ijk(gh)} - \bar{Y})^2
\]

\[
= n \sum_{h=1}^{m_{2}} (\bar{Y}_{h} - \bar{Y})^2 + n \sum_{g=1}^{m_{1}} (\bar{Y}_{gh} - \bar{Y}_{g})(\bar{Y}_{gh} - \bar{Y}_{h}) + \sum_{i=1}^{n} \sum_{j=1}^{n} (\bar{Y}_{ijk(gh)} - \bar{Y})^2
\]
Thus we have:

Total Sum of Squares = Sum of Squares between row means within Horizontal grids + Sum of Squares between Horizontal Grid Means + Sum of Squares between column Means within Vertical Grids + Sum of Squares between vertical grid Means + Sum of Squares between block means + Sum of Squares between treatment means + Error Sum of Squares.

Total degrees of freedom

\[ = n^2 - 1 = (m_2 - 1) + (n - m_2) + (m_1 - 1) + (n - m_1) + (n - 1) + (m_1 - 1)(m_2 - 1) + [(n - 1)(n - 2) - (m_1 - 1)(m_2 - 1)] \]

Total degrees of freedom = degrees of freedom for Horizontal Grids + degrees of freedom for rows within Horizontal Grids + degrees of freedom for Vertical Grids + degrees of freedom for columns within Vertical Grids + degrees of freedom for blocks + degrees of freedom for treatment + Error degrees of freedom.

\[
Total \ SS = \sum_{i=1}^{n} \sum_{j=1}^{n} \left( \sum_{g=1}^{m_2} \sum_{h=1}^{m_1} \left( y_{ijk} - \bar{Y} \right)^2 \right) = \sum_{i=1}^{n} \sum_{j=1}^{n} (y_{ijk} - \bar{Y})^2
\]

\[
SSR = \text{Row } SS = \sum_{i=1}^{n} \left( \sum_{j=1}^{n} \left( \bar{Y}_i - \bar{Y} \right)^2 \right) = \frac{1}{n} \sum_{i=1}^{n} R_i^2 - C.F., \text{ with } n - 1 \text{ d.f}
\]

\[
SSC = \text{Column } SS = \sum_{j=1}^{n} \left( \sum_{i=1}^{n} \left( \bar{Y}_j - \bar{Y} \right)^2 \right) = \frac{1}{n} \sum_{j=1}^{n} C_j^2 - C.F., \text{ with } n - 1 \text{ d.f}
\]

\[
SST = \text{Treatment } SS = \sum_{k=1}^{n} \left( \sum_{i=1}^{n} \left( \bar{Y}_k - \bar{Y} \right)^2 \right) = \frac{1}{n} \sum_{k=1}^{n} T_k^2 - C.F., \text{ with } n - 1 \text{ d.f}
\]

SS between Horizontal Grids = \[ \frac{\sum_{g=1}^{m_2} T_g^2}{nm_2} - \frac{G^2}{n^2} \]

SS between Vertical Grids = \[ \frac{\sum_{h=1}^{m_1} T_h^2}{nm_1} - \frac{G^2}{n^2} \]

Note that: \[ n \sum_{g=1}^{m_2} (\bar{Y}_g - \bar{Y})^2 + n \sum_{g=1}^{m_2} \sum_{i=1}^{m_1} (\bar{Y} - \bar{Y}_g)^2 = n \sum_{i=1}^{n} (\bar{Y}_i - \bar{Y})^2 \]

SS between Horizontal Grids + SS between row means within Horizontal Grids = SS between row means

Also: \( m_2 - 1 \) + \( n - m_2 \) = \( n - 1 \)

i.e. (degrees of freedom for Horizontal Grids) + (degrees of freedom for rows within Horizontal Grids)
\[ = \text{(degrees of freedom for rows)} \]

Similarly:
\[ n \sum_{h=1}^{m_1} \left( \bar{Y}_{jh} - \bar{Y} \right)^2 + n \sum_{h=1}^{m_1} \sum_{j=(h-1)m_2+1}^{hm_2} \left( \bar{Y} - \bar{Y}_{jh} \right)^2 = n \sum_{j=1}^{n} \left( \bar{Y}_j - \bar{Y} \right)^2 \]

SS between Vertical Grids + SS between column means within vertical Grids = SS between column means

Also: \( (m_1 - 1) + (n - m_1) = (n - 1) \)
i.e. (degrees of freedom for vertical Grids) + (degrees of freedom for columns within vertical Grids)

= (degrees of freedom for columns)

The layout and the structure of SSD show that the blocks are not orthogonal to rows and columns. In fact they are the joint effect of the Horizontal Grids and the Vertical Grids. Hence the SS between the Block Means is the interaction SS between the Horizontal Grids and the Vertical Grids.

\[
\text{SSB} = n^2 \sum_{g=1}^{m_2} \sum_{h=1}^{m_1} (\bar{Y}_{gh} - \bar{Y}_g - \bar{Y}_{.,h} + \bar{Y})^2
\]

\[
= \frac{\sum_{g=1}^{m_2} \sum_{h=1}^{m_1} T_{gh}^2}{n} - \frac{G^2}{n^2} - \frac{\sum_{g=1}^{m_2} T_{g.}^2}{nm_2} - \frac{G^2}{n^2} - \frac{\sum_{h=1}^{m_1} T_{.,h}^2}{nm_1} - \frac{G^2}{n^2}
\]

Hence SSB =
\[
\frac{\sum_{g=1}^{m_2} \sum_{h=1}^{m_1} T_{gh}^2}{n} - \frac{\sum_{g=1}^{m_2} T_{g.}^2}{nm_2} - \frac{\sum_{h=1}^{m_1} T_{.,h}^2}{nm_1} + \frac{G^2}{n^2}
\]

With \( n - m_1 - m_2 + 1 = (m_1 - 1)(m_2 - 1) \) number of degrees of freedom. \[^2\]

The Error SS is worked out by subtraction.

We arrive at the following Analysis of Variance table:
Sources of Variation | Degrees of freedom | SS | MS |
---|---|---|---|
**Rows**<br>Horizontal Grids | (m₂ – 1) | | |
Rows within Horizontal Grids | n – m₂ = m₂(m₁ – 1) | | |
Total | n – 1 | SSR | $s_R^2 = \frac{SSR}{n-1}$ |
**Columns**<br>Vertical Grids | (m₁ – 1) | | |
Cols. within Vertical Grids | n – m₁ = m₁(m₂ – 1) | | |
Total | n – 1 | SSC | $s_C^2 = \frac{SSC}{n-1}$ |
**Blocks**<br>Blocks | (m₁ – 1)(m₂ – 1) | SSB | $s_B^2 = \frac{SSB}{(m_1 - 1)(m_2 - 1)}$ |
**Treatments**<br>Treatments | n – 1 | SST | $s_T^2 = \frac{SST}{n-1}$ |
**Error**<br>Error | (n–1)(n–2) – (m₁–1) – (m₂–1) | SSE | $s_E^2 = \frac{SSE}{(n-1)(n-2) - (m_1 - 1)(m_2 - 1)}$ |
TOTAL | n² – 1 | TSS | |

**MOTE CARLO SIMULATIONS**

Let, in a forestry (agronomy, agriculture, etc.) experiment, we have four levels of nitrogen, designated by rows, four levels of potassium designated by columns, four types of soil variation designated by blocks and four varieties of a crop designated by treatments, A, B, C and D. We run a 4×4 Sudoku Square Design. Observations in the following table are pseudo random numbers:

No: 1

<table>
<thead>
<tr>
<th>Horizontal Grids</th>
<th>Col. Row</th>
<th>Vertical Grids</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>A(1)</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>D(3)</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>B(3)</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>C(9)</td>
</tr>
</tbody>
</table>
Here \( m_1 = 2, m_2 = 2 \) and \( n = 4 \)

We have: \( T_{11} = 10, \ T_{12} = 20, \ T_{1.} = 30 \)

\[ \sum_{g=1}^{m_2} T_{g.}^2 = 5800 \]

\( T_{21} = 30, \ T_{22} = 40, \ T_{2.} = 70 \)

\[ \sum_{h=1}^{m_1} T_{.h}^2 = 5200 \]

\( T_1 = 40, \ T_2 = 60, \)

\[ \sum_{g=1}^{m_1} \sum_{h=1}^{m_2} T_{gh}^2 = 3000 \]

By usual calculations we find:

Total SS = 275, SSR = 216, SSC = 51, SSB = 125, SST = 4

SSR + SSC + SSB + SST = 396, which exceeds Total SS, resulting in a negative Error SS. (Showing that the method of calculating SSB like that in Latin Square Design is not correct.)

Thus we use the method given in (1) above to find correct SSB.

\[
SS \text{ bet. HG} = \left( \sum_{g=1}^{m_2} T_{g.}^2 \right) / \left( nm_2 \right) - \left( G^2 / n^2 \right) = 100,
\]

\[
SS \text{ bet. VG} = \left( \sum_{h=1}^{m_1} T_{.h}^2 \right) / \left( nm_1 \right) - \left( G^2 / n^2 \right) = 25,
\]

SS for rows within HG = 216 – 100 = 116, SS for Cols. Within VG = 51 – 25 = 26

SSB = 3000/4 – 5800/8 – 5200/8 + 625 = 0

From this simulation, we come across the following conclusions:

(1) Since SSB = 0, therefore Sudoku Design is not appropriate for this type of situation. Loss of degrees of freedom will only over estimate the error mean square.

(2) SSB = 125, obtained by usual formula is actually made up of SS bet. HG (=100, which is included in SSR) and SS bet. VG (=25, which is included in SSC).

<table>
<thead>
<tr>
<th>No: 2</th>
<th>Vertical Grids</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Col. Row 1</td>
</tr>
<tr>
<td>Horizontal Grids</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>4</td>
</tr>
</tbody>
</table>
Here \( m_1 = 2, m_2 = 2 \) and \( n = 4 \)

We have: \( T_{11} = 30, \ T_{12} = 20, \ T_{1.} = 50 \sum_{g=1}^{m_2} T_{g.}^2 = 5000, \)

\( T_{21} = 10, \ T_{22} = 40, \ T_{2.} = 50 \sum_{h=1}^{m_1} T_{.h}^2 = 5200 \)

\( T_1 = 40, \ T_2 = 60, \ \sum_{g=1}^{m_2} \sum_{h=1}^{m_1} T_{gh}^2 = 3000 \)

By usual calculations we find:

Total SS = 275, SSR = 20, SSC = 83, SSB = 125, SST = 48

SSR + SSC + SSB + SST = 276, which exceeds Total SS, resulting in a negative Error SS. (Showing that the method of calculating SSB like that in Latin Square Design is not correct.)

Thus we use the method given in (1) above to find correct SSB.

\[
\text{SS bet. HG} = \left( \frac{\sum_{g=1}^{m_2} T_{g.}^2}{nm_2} \right) - \frac{G^2}{n^2} = 0, \quad \text{SS bet. VG} = \left( \frac{\sum_{h=1}^{m_1} T_{.h}^2}{nm_1} \right) - \frac{G^2}{n^2} = 50,
\]

SS for rows within HG = 20 – 0 = 20, SS for Cols. Within VG = 83 – 50 = 33

SSB = 3000/4 – 5000/8 – 5200/8 + 625 = 100

We arrive at the following analysis of variance:

<table>
<thead>
<tr>
<th>ANOVA Table</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sources of variation</td>
</tr>
<tr>
<td>---</td>
</tr>
<tr>
<td>Rows</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Columns</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Blocks</td>
</tr>
<tr>
<td>Treatments</td>
</tr>
<tr>
<td>Error</td>
</tr>
<tr>
<td>TOTAL</td>
</tr>
</tbody>
</table>
What we conclude from this simulation is that since there is a significant variation between block means therefore the appropriate design for this type of situation is Sudoku Square Design rather than Latin Square Design. Otherwise the error mean square will be highly over estimated, making our results insignificant.

DISCUSSION

The chief advantage of the Sudoku Square Design over the Latin Square Design is that it attempts to control the error by identifying a portion of the total variation with block means. The design is useful in the fields of agronomy, agriculture and forestry. A new blocking factor (blocks) is introduced to tackle the problem arising due to soil variation in the experimental area. Four different factors, each at n levels, identified by rows, columns, blocks and treatments can be tested simultaneously in one experiment. The blocks as blocking factor in Sudoku Square Design are only effective if the variance between blocks is significantly larger larger than the variance within blocks, otherwise the appropriate design may be the Latin Square Design, because unnecessary loss of degrees of freedom will increase the error mean square and make the results insignificant.

REFERENCES

Statistical meta-analysis for ordinal categorical data

Md Belal Hossain¹ and Shahjahan Khan
Department of Mathematics and Computing
Australian Centre for Sustainable Catchments
University of Southern Queensland
Toowoomba, AUSTRALIA.
Email: bjgaardar2003@yahoo.com and khans@usq.edu.au

ABSTRACT

Traditionally the odds ratio (OR) is used for measuring the extent of association between exposure and its binary outcomes in randomised controlled trials (RCTs) and similar studies. It is inapplicable if the outcomes are on an ordinal scale with more than two categories. In those studies, the generalised odds ratio (GOR) is used for summarising the difference between two stochastically ordered distributions of an ordinal categorical variable. Meta-analysis combines data from various independent trials in estimating the overall effect measure to make the sample size larger so that the inference based on the combined data is more reliable. In this paper we developed a method of meta-analysis using the GOR under independent multinomial sampling scheme for ordinal categorical data.

Keywords: Generalised odds ratio; Ordinal data; Multinomial distribution; Meta-analysis

1 Introduction

The first meta-analysis was performed by Pearson (1904) to overcome the problem of reduced statistical power in studies with small sample sizes; analyzing the results from a group of studies can allow more accurate data analysis [14]. Although meta-analysis is widely used in epidemiology and evidence-based medicine today, a meta-analysis of a medical treatment was not published until 1955. Glass (1976) and Hunter and Schmidt (1990) introduced more sophisticated analytical techniques in educational research in the 1970s. The method for meta-analysis introduced by Peto (1987) was the most widely used technique for adding together homogeneous studies. Then Thompson and Pocock (1987) concluded that meta-analysis provides clear qualitative conclusions but quantitative results have to be interpreted

¹On leave from Department of Statistics, University of Dhaka, Bangladesh
carefully as it fails to provide conclusive results for broad treatment policies when there exists heterogeneity among the studies.

It is common to use the relative risk (RR) or odds ration (OR) for measuring effect sizes for binary outcomes in epidemiology. But there are situations in which subjects are classified into several categories of severity of disease and exposures making RR and OR inappropriate. There are some measures proposed for handling outcomes of trials with more than two categories. [4] introduced some odds ratio statistics for the analysis of ordered categorical data assuming a specific model assumption. Later Clayton (1976) generalised the estimators for the case in which some observations are subject to censorship. McCullagh (1977) used paired comparisons on the ordinal variable. Whitehead and Jones (1994) developed stratified proportional odds model for ordinal outcomes transforming the \((j × l; j = 1, 2, \ldots, J; l = 1, 2, \ldots, L)\) contingency table for the ith (for \(i = 1, 2, \ldots, k\)) study into different combinations of \(2 × 2\) tables using log odds ratio as the effect measure. Whitehead et al. (2001) also proposed a proportional odds model on individual patient data by the log odds ratio with a general framework for fixed and random effects models. The above studies used either a \(2 × 2\) contingency table or a specific model to produce the treatment effect. Edwards and Baltzan (2000) proposed pooling \(\gamma\)’s instead of general odds ratio \(\text{OR}_C = (1 + \gamma)/(1 - \gamma)\), where \(\gamma\) is known as Goodman and Kruskal’s \(\gamma\) which is the same as Agresti (1980)’s \(\alpha\) using sample size weights although weighting by sample size may be misleading when heterogeneity exists among the studies. More importantly a variance estimate is a must for the pooled confidence interval (CI) of the \(\text{OR}_C\). Whereas pooling by inverse variance weighted method is more appropriate and makes sense in the presence of heterogeneity.

The generalised odds ratio (GOR) introduced by Agresti (1980) can be easily used as an effect measure for ordinal categorical data for the general \(J × L\) table. This measure is free from any specific model assumption and can be computed directly from the \(J × L\) table. This study attempts to employ GOR to measure the treatment effects for ordered categorical data and then combining the outcome using inverse variance weighted method in meta-analysis without merging the columns of \(J × L\) table. Pooling and using a common study effect for homogeneous studies in meta-analysis has an statistical agreement that in the forest plot the vertical line through the point estimate of the common pooled study effect passes through all the confidence intervals (CIs) of the individual study effects. Unfortunately in many studies there exist statistical disagreement to use the common pooled study effect if the outcome variable is heterogeneous. “Doing a meta-analysis is easy”, says Ingram Olkin but “Doing one well is hard” as heterogeneity among studies may lead to incorrect meta-analysis Mann (1990). For discussions on simple pooling and meta-analysis see Bravata and Olkin (2001). Further details on the topic can be found in Emerson (1994) and Egger and Smith (1997).
2 Generalized Odds Ratio

Let \( J \) be the number of comparison groups with \( L \) ordered outcome categories in each group. Then the \( J \times L \) contingency table represents the joint distribution of two ordinal categorical variables. In Table 1, \( X_{ij} \) is the count of the \( i \)th category in the \( j \)th group for the \( i \)th study, \( n_{ij} \) is the total count of \( j \)th group for the \( i \)th study, \( X_{11L} = n_{11} - X_{112} - X_{113} - \cdots - X_{11L} \), \( X_{12L} = n_{12} - X_{121} - X_{122} - \cdots - X_{12L} \), and \( X_{ijL} = n_{ij} - X_{ij1} - X_{ij2} - \cdots - X_{ijL} \).

When both \( J, L = 2 \), the GOR reduces to the OR for a single \( 2 \times 2 \) contingency table. In this study, we refer \( J=2 \) comparison groups in RCTs, namely the treatment group and the control group. More details about GOR and its mathematical formulation under independent multinomial sampling for randomized controlled trials are given below.

<table>
<thead>
<tr>
<th>Groups</th>
<th>Category 1</th>
<th>Category 2</th>
<th>\cdots</th>
<th>Category ( L )</th>
<th>Sample size</th>
</tr>
</thead>
<tbody>
<tr>
<td>Treatment</td>
<td>( X_{11} )</td>
<td>( X_{12} )</td>
<td>\cdots</td>
<td>( X_{1L} )</td>
<td>( n_{11} )</td>
</tr>
<tr>
<td>Control</td>
<td>( X_{21} )</td>
<td>( X_{22} )</td>
<td>\cdots</td>
<td>( X_{2L} )</td>
<td>( n_{12} )</td>
</tr>
</tbody>
</table>

The GOR is defined as the ratio of the two proportions of concordant and discordant pairs (Agresti 1980, 1990). A pair is said to be concordant if the subject ranked higher on groups/rows also ranks higher on categories/columns. Without loss of generality we assume that the response in category \( l' \) is more severe than the response in category \( l \) where \( l < l' \). Mathematically, the GOR for the \( i \)th is defined as

\[
\Gamma_i = (\Pi_d)^{-1}\Pi_c,
\]

where \( \Pi_c = \sum_{r=1}^{L-1} \sum_{s=r+1}^{L} \Pi_{r|1}\Pi_{s|2} \) and \( \Pi_d = \sum_{r=2}^{L} \sum_{s=1}^{r-1} \Pi_{r|1}\Pi_{s|2} \). Here, \( \Pi_c \) denotes the probability that the response of a randomly selected subject from group 2 is severer than the response of a randomly selected subject from group 1. Similarly, \( \Pi_d \) denotes the probability that the response of a randomly selected subject from group 1 is severer than the response of a randomly selected subject from group 2. The data with zero cell count is analysed adding \( \frac{1}{2} \) to each entry before calculation of the GOR. The value of \( \Gamma_i \) may vary from 0 to \( \infty \). \( \Gamma_i = 1 \) represents identical comparison groups as it is in odds ratio.

Suppose an independent random sample of size \( n_{ij} \) are taken from group \( j \) (\( j = 1, 2 \)) and \( X_{ijl} \) denote the count falling into category \( l \) of the \( i \)th study. Then the random vector \((X_{i1j}, X_{i2j}, \ldots, X_{iLj})\) follows the multinomial distribution with parameters \( n_{ij} \) and \( \pi_{ij} = (\pi_{i1j}, \pi_{i2j}, \ldots, \pi_{iLj}) \), where \( \pi_{ijl} \) is the probability of a subject to be in the \( l \)th category within the \( j \)th comparison group for the \( i \)th study.

The maximum likelihood estimator (MLE) of \( \pi_{iilj} \) is given by \( \hat{\pi}_{ilj} = X_{ilj}/n_{ij} \), for the \( i \)th study. For large \( n_{ij} \), \( \sqrt{n_{ij}}(\hat{\pi}_{ij} - \pi_{ij}) \), where \( \pi_{ij} = (\pi_{i1l}, \pi_{i2l}, \ldots, \pi_{iLl}) \), asymptotically
follows the L-dimensional multivariate normal distribution with mean vector \( \mathbf{0} \) and \( L \times L \) covariance matrix with the diagonal entries \( \pi_{ij}(1 - \pi_{ij}) \), and off-diagonal entries \(-\pi_{il} \pi_{lj} \) for \( l \neq l' \). The MLE of \( \Gamma_i \), say \( \hat{\Gamma}_i \), is defined as
\[
\hat{\Gamma}_i = (\hat{\Pi}_d)^{-1}\hat{\Pi}_c
\]
where \( \hat{\Pi}_c = \sum_{s=1}^{L-1} \sum_{r=s+1}^{L} \hat{\pi}_{r|s} \hat{\pi}_{s|2} \) and \( \hat{\Pi}_d = \sum_{r=2}^{L} \sum_{s=1}^{r-1} \hat{\pi}_{r|1} \hat{\pi}_{s|2} \) for the \( i \)th study. In addition to the MLE of \( \Gamma_i \), the variance of the MLE is required to construct the CI for \( \Gamma_i \). Using the delta method [2], the asymptotic variance of \( \hat{\Gamma}_i \) becomes
\[
\text{Asy.Var}(\hat{\Gamma}_i) = \sum_{r=1}^{L} \frac{\sum_{s=1}^{r-1} \pi_{s|2} \pi_{r|1}}{n_i \Pi_d^2} + \sum_{s=1}^{L} \frac{\sum_{r=s+1}^{L} \pi_{r|1} \pi_{s|2}}{n_2 \Pi_d^2}.
\]

Here by convention \( \sum_{l=1}^{L} \pi_{lj} = 0 \) and \( \sum_{l=1}^{0} \pi_{lj} = 0 \) for \( l=1, 2 \). To estimate \( \text{Var}(\hat{\Gamma}_i) \), say \( \hat{\text{var}}(\hat{\Gamma}_i) \), we substitute \( \hat{\pi}_{ilj} \) for \( \pi_{ilj} \), \( \hat{\Gamma}_i \) for \( \Gamma_i \), and \( \hat{\Pi}_d \) for \( \Pi_d \) in the definition of \( \hat{\Gamma}_i \).

For large \( n_{ij} \), an asymptotic \( 100(1 - \alpha) \) percent confidence interval for the \( \Gamma_i \) is given by
\[
\left[ \max\left\{ \hat{\Gamma}_i - Z_{\alpha/2} \sqrt{\hat{\text{var}}(\hat{\Gamma}_i)}, 0 \right\}, \hat{\Gamma}_i + Z_{\alpha/2} \sqrt{\hat{\text{var}}(\hat{\Gamma}_i)} \right],
\]
where \( Z_{\alpha/2} \) is the upper \( (100 - \frac{\alpha}{2}) \)th percentile of the standard normal distribution.

### 2.1 Meta-analysis for fixed effects model

In the fixed effect model the effect the outcome variables are assumed to be drawn from the same population. For \( i = 1, 2, \cdots, k \) independent studies if \( \hat{\Gamma}_i \) represents logarithm of GOR, the observed effect size with variance \( \nu_i \), then assuming \( \Gamma_1 = \Gamma_2 = \cdots, \Gamma_k = \Gamma_0 \), a pooled estimate of the treatment effect is given by
\[
\hat{\Gamma}_0 = \frac{\sum_i \omega_i \hat{\Gamma}_i}{\sum_i \omega_i}.
\]
For an arbitrary number of outcome categories \( (L) \) in RCTs in which each row is modeled as an independent multinomial distribution, the estimated variance of the \( i \)th study is
\[
\hat{\Omega}_i^{-1} = \sum_{l=1}^{L-1} \sum_{j=1}^{2} \frac{1}{n_{ij} \hat{\pi}_{ijl}(1 - \hat{\pi}_{ijl})}
\]
where \( n_{ij} \) is the total count of the \( j \)th group for the \( i \)th study, \( \hat{\pi}_{ijl} = X_{ijl}/n_{ij} \) is the MLE of \( \pi_{ijl} \) and \( X_{ijl} \) is the count of the \( l \)th category in the \( j \)th group for the \( i \)th study.
Assuming \( \hat{\Gamma}_i \)'s are normally distributed, an approximate 100(1 - \( \alpha \))% CI for the \( i \)th GOR is given by the formula

\[
\exp \left[ \hat{\Gamma}_i \pm z_{\alpha/2} \omega_i^{-1/2} \right].
\]  

(7)

If \( \hat{\Gamma}_0 \) is assumed to be normally distributed, an approximate 100(1 - \( \alpha \))% confidence interval (CI) for the population effect, \( \Gamma_0 \), is given by

\[
\exp \left[ \hat{\Gamma}_0 \pm z_{\alpha/2} \omega^{-1/2} \right],
\]  

(8)

for the meta-analysis where \( \omega^{-1} = \text{var}(\hat{\Gamma}_0) = 1 / \sum_{i=1}^{k} \omega_i \) and \( z_{\alpha/2} \) is the \( \alpha/2 \times 100 \) percentage point of a standard normal distribution.

### 2.2 Meta-analysis for random effects model

In the random effects model the outcome variables are considered to be randomly drawn from different populations. Define \( \omega \) and \( s_{W^2} \) to be the mean and variance of the weights from the \( k \) studies:

\[
\omega = \frac{1}{k} \sum_{i=1}^{k} \omega_i \quad \text{and} \quad s_{W^2} = \frac{1}{k-1} \left( \sum_{i=1}^{k} \omega_i^2 - k \omega^2 \right).
\]

(9)

Further, define

\[
U = (k - 1) \left( \omega - \frac{s_{W^2}}{\omega} \right) \quad \text{and} \quad Q = \sum_{i=1}^{k} \omega_i (\hat{\Gamma}_i - \Gamma_l)^2,
\]

(10)

where \( Q \) is the heterogeneity test statistic, also known as Cochran's \( \chi^2 \) statistic [?] for testing the \( H_0 = \Gamma_1 = \Gamma_2 = \cdots = \Gamma_k = \Gamma_0 \). The estimated component of variance due to inter-study variation in effect size, \( \hat{\tau}^2 \), is calculated as

\[
\hat{\tau}^2 = \begin{cases} 
0 & \text{if } Q \leq (k - 1)/U \\
(Q - (k - 1))/U & \text{if } Q > (k - 1).
\end{cases}
\]

(11)

A 100(1 - \( \alpha \))% CI for \( \Gamma_l \), is given by

\[
\exp \left[ \hat{\Gamma}_{iR} \pm z_{\alpha/2} / \sqrt{\omega^2_{iR}} \right],
\]

(12)

where \( \omega^2_{iR} = \frac{1}{1/\omega_i^2 + \tau^2} \), under the assumption of normality.

The point estimate for the mean treatment effect of all studies, \( \Gamma_0 \), can be computed by

\[
\hat{\Gamma}_{0R} = \frac{1}{k} \sum_{i=1}^{k} \omega_i^2 \hat{\Gamma}_i / \sum_{i=1}^{k} \omega_i^2.
\]

(13)

If normality of \( \Gamma_{0R} \) is assumed, a 100(1 - \( \alpha \))% CI for \( \Gamma_0 \) is given by

\[
\exp \left[ \hat{\Gamma}_{0R} \pm z_{\alpha/2} / \sqrt{\sum_{i=1}^{k} \omega_i^2} \right].
\]

(14)
3 Conclusions

In this paper, we develop a meta-analysis method using GOR for multi-levels ordinal categorical outcomes. The currently available proportional odds model is restricted to the proportionality assumption and there is no well defined variance estimate of the pooled estimate for the sample size weight method. Use of the LOR or similar effect measures for multi-levels ordinal outcomes by collapsing a $2 \times L$ table into $2 \times 2$ tables causes loss of information, inflate the estimate and inappropriately reduce the spread.

The proposed meta-analysis method using GOR is very simple and has straightforward interpretation. It has simple variance estimate for individual study and meta-analysis. It can also be used for binary and latent continuous outcomes. Therefore, GOR is a very useful and superior effect measure in meta-analysis for the ordinal categorical data.

Acknowledgement

The authors are grateful to Australian Centre for Sustainable Catchments, University of Southern Queensland, Australia and University of Dhaka, Bangladesh for sponsoring this work.

References


FACTOR ANALYSIS OF TEACHER’S INTERVENTION STRATEGIES FOR SOLVING STUDENT’S BEHAVIOR AND LEARNING PROBLEMS
(A Cross-Sectional Case Study based on the Teachers of Girls Schools of Ravi Town, Lahore)

Sarah Ashraf and Asifa Kamal
Department of Statistics, Lahore College for Women University, Lahore, Pakistan
Email: asifa.k53@gmail.com

ABSTRACT

The purpose of this research was to assess the teacher’s intervention strategies for solving student’s behavior and learning problems. Three types of behavior problems were taken into account i.e. behavior problems related to immature child, child with perceptual difficulty and unmanageable child. A sample of 290 teachers was selected from both government and private girls’ schools. Data collection was done through a questionnaire which comprises of 39 possible interventions. Descriptive analysis and multivariate analysis were carried out with the help of the software SPSS. Seven Factors were identified after applying factor analysis techniques using orthogonal rotation method. These factors were labeled as family related intervention, negative intervention, teacher’s classroom behavior intervention, social intervention, performance based intervention, consultative actions interventions and interventions related to teaching style. The overall mean rating of importance was higher for teacher’s classroom behavior intervention, teaching style intervention, social intervention; interventions related to consultative actions and performance based intervention. Family related interventions were given relatively lower importance. The negative interventions were given least importance by the teachers. Though teachers were very well aware of interventions nature of interventions but the results showed that they would use the same interventions strategies for three types of behavioral problems.

KEYWORDS
Interventions; principal component factor analysis; immature child; Child with perceptual difficulty; Unmanageable child; Kruskal Wallis.

1. INTRODUCTION

Development of any nation is impossible without education. Schools are the platform to start education in modern world. Education has not been given due importance and in recent years more emphasis has been paid to higher education in Pakistan. While without promoting school education how can we attain our aims of higher education. Many countries are facing the problem of drop outs in their schools. School environment and lack of teacher’s knowledge about behavioral issues of children are major causes for this. If behavior problems are handled psychologically then students drop out rate can be reduced.
Behavioral disorders or conduct disorders are increasing dramatically in our classrooms. As a result their presence severely constrains the ability of the school systems to educate students effectively. There are three types of behavior problem among students which generally teachers face at school i.e. immature children, children with perceptual difficulties and unmanageable children (Algozzine et al. 1983).

The first type of children usually does not act according to the chronological age. This type of children remain in a complex that they are physically not as good as the other children. They lack confidence in their conversation and their behavior is unpredictable. Teacher can handle such child by indulging him in different activities and also by paying more attention. Involvement of parents also helps to resolve the problem.

The second type of behavioral problem discussed in this paper is a child with perceptual difficulty. Children like this find difficulty in understanding the alphabets and resultantly can not accurately read or write. They cannot assess body distance and also face accidents. Parents should take part to resolve this type of behavioral problem. Consultation with doctors is first and foremost for mental checkup of the child.

Third type consists of rude, rebellious and restless children. In simple words they cannot be easily managed. These children are hard to control and show arrogant attitude towards everyone. Punishment is not the solution for this type of behavioral problem, rather it aggravates the situation. Parents should express love, affection and no doubt attention towards such child. Teacher should also engage the child in new hobbies and activities etc. Webster-Stratton et al. (2008) has compared intervention teachers with control teachers. He concluded that intervention teachers used more positive classroom management approach and their students showed more social capability and emotional self control than control teachers and students. Intervention teachers showed more involvement with parents than control teachers. Epstein et al. (2008) has given a guideline for effective intervention strategies that promote the positive student behavior such as recognition of problem, new classroom atmosphere to reduce the problem, introduction of new expertise which help positive behavior in class room, involvement of family and other teachers for help, assessment of behavior problems in school and its control strategies. Effect of in-service training of teachers for behavioral problem of children with attention-deficit / hyperactivity disorder (ADHD) was studied by Jones et al. (2008). Algozzine et al. (1983) conducted a case study in which 174 elementary school teachers were involved. They were given the summary of third grade male student who actually has unmanageable behavior, immature behavior or perceptual difficulties in the classroom. The least favored interventions were the non-teacher-directed actions. Different kinds of intervention were recommended for different kinds of behavior. The drug therapy which is considered as severe intervention was recommended for unmanageable student. Peer tutoring which is considered as less severe intervention was recommended for student with perceptual difficulties.

2. OBJECTIVE OF THE STUDY

The proposed study would analyze the teacher’s opinion in the form of intervention to handle the behavioral problem of students and suggest policy makers to take actions for teacher trainings so that they can handle behavioral problems of students which cause hindrance in their learning.
3. DATA AND METHODS

Data was collected from the teaching staff of schools of one union council of Lahore. The respondents of proposed study were teachers from both Government and Private Schools for girls. Stratified random sampling has been used to choose sample from government and private schools. Questionnaire on five point likert scale was designed for 3 types of behavioral problems of students separately following the Algozzine et al. (1983) i.e. Case Study I for immature child, Case Study II for child with perceptual difficulty and Case Study III for unmanageable child. The value of cronbach’s alpha for questionnaire was 0.880 which means questionnaire designed was reliable to serve the purpose. Factor analysis technique and Kruskal Wallis was used for statistical analysis.

4. RESULTS

4.1 Descriptive Analysis

Number of students in class is a very important factor. If there is a large number of students than the teacher will not be able to pay individual attention. Resultantly he or she cannot resolve the general and more specifically the behavioral problems of students. The number of students in each section was generally between 20 and 50 but the percentage of teachers having number of students above 50, was also quite high (Table 1).

<table>
<thead>
<tr>
<th>Information collected</th>
<th>Percentages</th>
<th>Information collected</th>
<th>Percentages</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. of students in class</td>
<td></td>
<td>No. of classes taught</td>
<td></td>
</tr>
<tr>
<td>Less than 25</td>
<td>14.8</td>
<td>Less than 4</td>
<td>7.9</td>
</tr>
<tr>
<td>25-50</td>
<td>63.1</td>
<td>4-8</td>
<td>89.3</td>
</tr>
<tr>
<td>Above 50</td>
<td>22.1</td>
<td>Greater than 8</td>
<td>2.8</td>
</tr>
<tr>
<td>Education</td>
<td></td>
<td>Experience of teacher</td>
<td></td>
</tr>
<tr>
<td>Matriculation</td>
<td>0.7</td>
<td>Less than 1 year experience</td>
<td>17.6</td>
</tr>
<tr>
<td>Intermediate</td>
<td>18.3</td>
<td>01-10</td>
<td>73.1</td>
</tr>
<tr>
<td>Bachelors</td>
<td>57.9</td>
<td>10-20</td>
<td>6.2</td>
</tr>
<tr>
<td>Masters</td>
<td>22.8</td>
<td>Above 20 years</td>
<td>3.1</td>
</tr>
<tr>
<td>M. Phil</td>
<td>0.3</td>
<td>Average Percentage result for past 3 years</td>
<td></td>
</tr>
<tr>
<td>Professional Education</td>
<td></td>
<td>60-69</td>
<td>5.2</td>
</tr>
<tr>
<td>No degree</td>
<td>46.0</td>
<td>70-79</td>
<td>48.3</td>
</tr>
<tr>
<td>CT/PTC</td>
<td>35.7</td>
<td>80-89</td>
<td>13.8</td>
</tr>
<tr>
<td>B.Ed</td>
<td>16.6</td>
<td>90-100</td>
<td>32.8</td>
</tr>
<tr>
<td>M.Ed</td>
<td>1.7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Studied child psychology</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>63.4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>36.6</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Maximum percentage of teachers in schools was graduates. Some of them were even M.Phil degree holder but only few of these teachers have any professional education. Most of the teachers have 4 to 8 classes per day. Percentage of teachers who have studied child psychology of school going children at any level of their education was very small.
Most of the teaching staff has teaching experience of 1 to 10 years. Hence the sample consists of teachers who have enough experience to understand the behavioral problems and their solutions.

The means and standard deviations (S.D.) of all the variables regarding teacher’s intervention were computed and shown in the table 2. The entries of this table give a clear picture of these variables using a single précised/centered value. The standard deviation of these variables explains the amount of variability in responses. A low standard deviation indicates that the response data points tend to be very close to the same value of mean, while high standard deviation indicates that the responses of these variables are spread out over a large range of values.

From table 2 it is observed that twenty four questions/ items have average rating 4.0 or greater, indicating the teacher’s agreement among themselves to use these interventions for dealing with students having different behavior problems in the class room. The most favored interventions are related to teaching style (Q1, Q2, Q3, Q4, Q5, Q9), social interventions (Q8, Q12, Q16, Q17), interventions related consulting actions (Q14), performance based interventions ( Q15, Q28, Q29, Q30, Q31, Q32), teacher’s class room behavior interventions(Q33, Q34, Q35, Q36, Q37, Q38, Q39). Least favored interventions are for (Q7, Q10, Q11, Q13, Q18) which are related to the factor negative interventions and average ratings for these interventions is less than two.

### Table 2: Descriptive Statistics for interventions

<table>
<thead>
<tr>
<th>Interventions</th>
<th>N</th>
<th>Mean</th>
<th>S. D.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q1: Meet with education experts</td>
<td>290</td>
<td>4.49</td>
<td>.534</td>
</tr>
<tr>
<td>Q2: Reading of text books and articles</td>
<td>290</td>
<td>4.48</td>
<td>.507</td>
</tr>
<tr>
<td>Q3: Use of special instructional material (computer etc)</td>
<td>290</td>
<td>4.40</td>
<td>.511</td>
</tr>
<tr>
<td>Q4: Prepare special home work</td>
<td>290</td>
<td>4.47</td>
<td>.520</td>
</tr>
<tr>
<td>Q5: Provide some special instructional material for home work</td>
<td>290</td>
<td>4.39</td>
<td>.543</td>
</tr>
<tr>
<td>Q6: Interview with family to detect problem</td>
<td>290</td>
<td>3.95</td>
<td>.615</td>
</tr>
<tr>
<td>Q7: Retain the child in the same class</td>
<td>290</td>
<td>1.94</td>
<td>.728</td>
</tr>
<tr>
<td>Q8: Arrange the company of good students</td>
<td>290</td>
<td>4.26</td>
<td>.639</td>
</tr>
<tr>
<td>Q9: Provide teacher services for private tutoring</td>
<td>290</td>
<td>4.34</td>
<td>.631</td>
</tr>
<tr>
<td>Q10: Place child in another class</td>
<td>290</td>
<td>1.89</td>
<td>.692</td>
</tr>
<tr>
<td>Q11: Show strict behavior towards that child</td>
<td>290</td>
<td>1.99</td>
<td>.778</td>
</tr>
<tr>
<td>Q12: Involve child’s friends</td>
<td>290</td>
<td>4.18</td>
<td>.684</td>
</tr>
<tr>
<td>Q13: Physically punish the child</td>
<td>290</td>
<td>1.98</td>
<td>.806</td>
</tr>
<tr>
<td>Q14: Consult principal to get his/her suggestion</td>
<td>290</td>
<td>4.09</td>
<td>.663</td>
</tr>
<tr>
<td>Q15: Provide feedback to principal for the performance of child</td>
<td>290</td>
<td>4.20</td>
<td>.597</td>
</tr>
<tr>
<td>Q16: Arrange social skills training</td>
<td>290</td>
<td>4.21</td>
<td>.661</td>
</tr>
<tr>
<td>Q17: Ask principal to provide facility for speech/language therapy</td>
<td>290</td>
<td>4.20</td>
<td>.629</td>
</tr>
<tr>
<td>Q18: Request the principal to expel the child</td>
<td>290</td>
<td>1.95</td>
<td>.781</td>
</tr>
<tr>
<td>Q19: Provide feedback to parents for up to date performance of child</td>
<td>290</td>
<td>3.92</td>
<td>.576</td>
</tr>
<tr>
<td>Q20: Suggest parents to consult medical doctor</td>
<td>290</td>
<td>3.97</td>
<td>.708</td>
</tr>
<tr>
<td>Q21: Refer the child to special education expert</td>
<td>290</td>
<td>3.98</td>
<td>.597</td>
</tr>
</tbody>
</table>
### Interventions

| Q22: Convince parents to spend more time with child | 290 | 3.94 | .456 |
| Q23: Ask parents to provide facility for speech/ language therapy | 290 | 3.92 | .517 |
| Q24: Suggest parents to control behavior with drug medication | 290 | 3.91 | .732 |
| Q25: Involve child’s brothers and sisters | 290 | 3.93 | .511 |
| Q26: Hold extra parents conferences for particular problematic child | 290 | 3.92 | .453 |
| Q27: Seek help from parents to resolve the problem | 290 | 3.92 | .471 |
| Q28: Periodically evaluate the performance of child after psychological treatment | 290 | 4.01 | .622 |
| Q29: Get the child’s test score in other subjects | 290 | 4.20 | .593 |
| Q30: Get the knowledge of his/her IQ test score | 290 | 4.21 | .525 |
| Q31: Obtain the knowledge of personality test result | 290 | 4.12 | .542 |
| Q32: Obtain the knowledge of speech and language test score | 290 | 4.13 | .532 |
| Q33: Keep the eye contact with children | 290 | 4.52 | .500 |
| Q34: Move around the class | 290 | 4.47 | .514 |
| Q35: Give permission of breaks during period | 290 | 4.50 | .541 |
| Q36: Give some creative and interesting ideas | 290 | 4.61 | .489 |
| Q37: Sometimes use humor in the class to lighten the burden of tough lesson | 290 | 4.52 | .534 |
| Q38: Approach that child in class room daily | 290 | 4.45 | .498 |
| Q39: Organize the break after some periods | 290 | 4.43 | .524 |

| Valid N (listwise) | 290 |

---

### 4.2 Multivariate Analysis

The teacher’s intervention strategies for solving student’s behavior and learning problems on these attributes are convenient, if these attitudes can be “grouped” or reduce these 39 variables to a smaller number. To identify factors of teacher’s intervention strategies SPSS was used. Before application of factor analysis technique assumption of sphericity was checked. The Bartlett’s test of Sphericity yielded a value of 7023.445 with associated level of significance smaller than 0.05 which indicates that sufficient correlations exist among the variables to proceed for factor analysis. Measure of sampling adequacy overall (Kaiser-Meyer-Olkin) is 0.864 which confirms the appropriateness of factor analysis technique for the data. The rotated component matrix was computed by using varimax rotation (Table 3). The factors thus obtained are given in the table 4.
### Table 3: Rotated Component Matrix

<table>
<thead>
<tr>
<th>Component</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.741</td>
<td></td>
</tr>
<tr>
<td>Q2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.581</td>
<td></td>
</tr>
<tr>
<td>Q3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.705</td>
<td></td>
</tr>
<tr>
<td>Q4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.658</td>
<td></td>
</tr>
<tr>
<td>Q5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.633</td>
<td></td>
</tr>
<tr>
<td>Q6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.624</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q7</td>
<td></td>
<td></td>
<td></td>
<td>.904</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.821</td>
<td></td>
</tr>
<tr>
<td>Q9</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.531</td>
</tr>
<tr>
<td>Q10</td>
<td></td>
<td></td>
<td>.818</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q11</td>
<td></td>
<td></td>
<td>.842</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q12</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.828</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q13</td>
<td></td>
<td></td>
<td>.915</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q14</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.819</td>
</tr>
<tr>
<td>Q15</td>
<td></td>
<td></td>
<td></td>
<td>.444</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q16</td>
<td></td>
<td></td>
<td></td>
<td>.893</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q17</td>
<td></td>
<td></td>
<td></td>
<td>.852</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q18</td>
<td></td>
<td></td>
<td></td>
<td>.953</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q19</td>
<td></td>
<td>.687</td>
<td></td>
<td></td>
<td></td>
<td>.835</td>
<td></td>
</tr>
<tr>
<td>Q20</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.858</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q21</td>
<td></td>
<td></td>
<td>.776</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q22</td>
<td></td>
<td></td>
<td>.794</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q23</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.868</td>
<td></td>
</tr>
<tr>
<td>Q24</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.866</td>
</tr>
<tr>
<td>Q25</td>
<td></td>
<td></td>
<td>.846</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q26</td>
<td></td>
<td></td>
<td>.866</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q27</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.819</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q28</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.749</td>
<td></td>
</tr>
<tr>
<td>Q29</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.667</td>
<td></td>
</tr>
<tr>
<td>Q30</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.732</td>
<td></td>
</tr>
<tr>
<td>Q31</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.730</td>
<td></td>
</tr>
<tr>
<td>Q32</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.743</td>
<td></td>
</tr>
<tr>
<td>Q33</td>
<td></td>
<td></td>
<td></td>
<td>.699</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q34</td>
<td></td>
<td></td>
<td></td>
<td>.683</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q35</td>
<td></td>
<td></td>
<td></td>
<td>.618</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q36</td>
<td></td>
<td></td>
<td></td>
<td>.653</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q37</td>
<td></td>
<td></td>
<td></td>
<td>.703</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q38</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.735</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Factor loading’s greater than 0.40 are considered in table 3.
Table 4:
Labeling of Factors

<table>
<thead>
<tr>
<th>Factors</th>
<th>Group</th>
<th>Name of Group</th>
</tr>
</thead>
<tbody>
<tr>
<td>F1</td>
<td>Q6, Q19, Q22, Q23, Q25, Q26, Q27</td>
<td>Family related intervention.</td>
</tr>
<tr>
<td>F2</td>
<td>Q7, Q10, Q11, Q13, Q18.</td>
<td>Negative intervention.</td>
</tr>
<tr>
<td>F3</td>
<td>Q33, Q34, Q35, Q36, Q37, Q38.</td>
<td>Teacher’s class room behavior intervention.</td>
</tr>
<tr>
<td>F4</td>
<td>Q12, Q16, Q17, Q8.</td>
<td>Social intervention.</td>
</tr>
<tr>
<td>F5</td>
<td>Q15, Q28, Q29, Q30, Q31, Q32.</td>
<td>Performance based intervention.</td>
</tr>
<tr>
<td>F6</td>
<td>Q14, Q20, Q21, Q24.</td>
<td>Interventions related to consultancy</td>
</tr>
<tr>
<td>F7</td>
<td>Q1, Q2, Q3, Q4, Q5, Q9.</td>
<td>Interventions related to teaching style.</td>
</tr>
</tbody>
</table>

4.3 Testing Equality of Average rating of three Case Studies
For comparing average ratings of three case studies (case studies for immature, unmanageable and students with perceptual difficulty) Kruskal wallis test was used.

Table 5:
Relative importance of intervention factors for teachers

<table>
<thead>
<tr>
<th>Intervention factors</th>
<th>Immature child case study1</th>
<th>Unmanageable child case study2</th>
<th>Child with perceptual difficulties Case study3</th>
<th>Mean rating of Importance</th>
<th>Kruskal Wallis p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Family related intervention (F1)</td>
<td>3.93(0.29)</td>
<td>3.91(0.438)</td>
<td>3.93(.442)</td>
<td>3.9233</td>
<td>0.543</td>
</tr>
<tr>
<td>Negative intervention (F2)</td>
<td>1.9794(.704)</td>
<td>1.9361(.677)</td>
<td>1.9354(.649)</td>
<td>1.9503</td>
<td>0.904</td>
</tr>
<tr>
<td>Teacher’s class room behavior intervention (F3)</td>
<td>4.5464(.391)</td>
<td>4.4811(.373)</td>
<td>4.5087(.369)</td>
<td>4.5120</td>
<td>0.517</td>
</tr>
<tr>
<td>Social intervention (F4)</td>
<td>4.2191(.598)</td>
<td>4.2139(.5694)</td>
<td>4.2083(.656)</td>
<td>4.2137</td>
<td>0.871</td>
</tr>
<tr>
<td>Performance based intervention. (F5)</td>
<td>4.1598(.418)</td>
<td>4.1117(.401)</td>
<td>4.1632(.437)</td>
<td>4.1449</td>
<td>0.455</td>
</tr>
<tr>
<td>Interventions related consulting actions. (F6)</td>
<td>4.0000(.571)</td>
<td>3.9897(.614)</td>
<td>3.9688(.585)</td>
<td>3.9861</td>
<td>0.900</td>
</tr>
<tr>
<td>Intervention related to teaching style. (F7)</td>
<td>4.4244(.369)</td>
<td>4.4210(.380)</td>
<td>4.4410(.397)</td>
<td>4.4288</td>
<td>0.962</td>
</tr>
</tbody>
</table>

Difference between average ratings is statistically insignificant for three case studies (immature, unmanageable and child with perceptual difficulties) for all the seven intervention factors (Table 5). These results depict that teachers used to follow the same intervention strategies for all three types of students. Also it is observed from the (Table 3) that most important interventions factors were 3,4,5,6 and 7 as mean rating is
high and these have similar ratings. Family intervention factor is important but its importance is less than the above other mentioned factors. Its mean rating is approximately “3” meaning respondent is neutral for this intervention factor to be opted or not. Negative interventions are found to be least important.

5. COMMENTS AND CONCLUSION

Teacher’s participation in solving student’s difficulties in school has a significant importance. Teachers were asked appropriate questions so that their intervention strategies for solving student’s behavior and learning problems could be identified. The interventions which are mostly preferred by the teachers are, to meet with experts of education, reading of textbook and articles, use of special instructional material (computer etc), prepare social home work, provide some special instructional material for homework, provides teacher services for private tutoring, arrange the company of good students, involve child’s friends, arrange social skills training, ask principal to provide facility for speech/ language therapy, consult principal to get his / her suggestions, provide feedback to principal for the performance of child, periodically evaluate the performance of child after psychological treatment, get child’s test score in other subjects, get the knowledge of his/her IQ test score, obtain the knowledge of personality test result, obtain the knowledge of speed and language test score, keep the eye contact with children, move around the class, give permission of breaks, give some creative and interesting ideas, sometimes use humor in the class to light the burden of tough lesson, approach the child particularly in class while teaching and organize the break when found child is tired. However the least favored interventions are to retain the child in the same class, to place child in another class, show strict behavior towards the child, physically punish the child and request the principal to expel the child.

To reduce or summarize the data, a factor analysis technique that would conveniently reduce data without much loss of information can be used. After applying factor analysis seven factors have been extracted from the data which give clear picture of intervention strategies of teachers. These factors were given appropriate names according to the type of intervention. The overall mean rating of importance is higher for teacher’s classroom behavior intervention, teaching style intervention, social intervention; interventions related to consultative actions and performance based intervention. Family related intervention has relatively lower importance as compared to the above factors. This depicts positive picture of teacher’s attitude towards understanding of behavioral problems of child and interventions opted to resolve it. It seems that teachers realize their role in solution of problem and rate those interventions high, which are related to teachers themselves or to school. They also realize the importance of involvement of parents in the solution of child’s problem. Negative interventions have been rejected by the teachers which is a very positive point.

Though teachers are very well aware of interventions but they deal the three types of behavioral problems in the same way. Importance of intervention strategies is the same for immature child, unmanageable child and for child with perceptual difficulties. This might be due to the reason that they lack the knowledge of psychology of school going children to handle the specific behavior problem. If the teachers are trained, the behavioral and learning problems of the students can be tackled appropriately.
REFERENCES


Wireless Networking technology is now one of the most popular technologies but still there are drawbacks which are closely associated with Wireless Networks. In wireless network communication, the data is transferred from one point to another point through radio waves which makes wireless networks weak for attacks. To eliminate threats, understanding about said attacks always provides good ability to defend wireless network. This research paper will describe the overview of the wireless technology with its drawbacks, Present security and privacy issues, Potential wireless network Security threats (unauthorized access, Active eavesdropping, Man in the Middle Attack, Denial of Service and etc), Tools that hackers often use to exploit vulnerabilities in wireless networking (NetStumbler, Kismet, Airsnort, Airsnarf, Airjack and etc) and will propose the best ways to secure Wireless Network.

1. INTRODUCTION

The home and business users are implementing wireless technology because of its convenience, easy installation, troubleshooting and low price in the market but security is the big concern for this technology because these devices can be attacked by the malicious attacks [1]. The Wireless networks must defend these attacks as illustrated in this paper.

Common wireless network vulnerabilities are as under:
- End users are not security experts, and may not be aware of the risks posed by wireless LANs.
- Approximately all access points having default configurations have not activated WEP security.
- Most of the users do not change access point’s default key used by all the vendor's products out of the box.
- The Wireless Access Points which are enabled with WEP can be cracked easily.

To access the internet in wireless network, the clients are connected with the Access Point and that AP is connected to the wireless router. The function of the wireless router is to broadcast a signal through the air and all the wireless clients within the range can connect to the wireless network.

IEEE has developed wireless network standard that is called 802.11. This standard unfortunately has limited support and still could not provide full security, privacy and confidentially through Wireless Equivalent Privacy (WEP) [2]. The development of 802.11 standards was started in late 1990s but true development has begun in 2000-2001.
This paper will describe security threats to IEEE 802.11 wireless network. The initial problem in the 802.11 is that it used media access control layer mechanism to talk with other networks.

2. WIRELESS EQUIVALENT PRIVACY (WEP)

IEEE 802.11 uses Wired Equivalent Privacy (WEP) for security purpose against eavesdropping and other attacks that are used to hack a wireless network but unfortunately, still there are number of flaws /loopholes in the protocol [3].

Wireless Equivalent Privacy (WEP) use RC4 algorithm and is based on data link layer security technology [3].

The first drawback of the WEP is that it uses secret user key that is called base key in RC4 algorithm. The main goal of the secret key is to encrypt the data over the wireless network and same time; Cyclic Redundancy Check (CRC) is used to protect the integrity of the packets [3].

WEP uses four different types of the base keys. In 64 bit encryption, only 40 bits are encrypted and the remaining 24 bits are system generated bits. This thing is big drawback of RC4 algorithm as it is very easy to crack 40 bits key. Similarly in 128 bits encryption, 104 bits are used for encryption not 128 and vice versa [4].

It is pertinent to mention here that both senders and receivers use same secret key.

To generate cipher text in RC4 cipher algorithm, the sender client computer XOR the secret key with plain text and similarly the receiver computer who have already same secret key XOR the cipher text to generate original sent text as depicted below [3].

\[
\begin{align*}
\text{Plaintext} & \quad \text{XOR} \quad \text{Message} \quad \text{CRC} \\
& \quad \text{Keystream} = \text{RC4}(v,k) \\
& \quad \text{Ciphertext} \\
\text{Transmitted Data} & \quad \text{V}
\end{align*}
\]

Fig. 1

The above stated working of RC4 cipher algorithm clearly shows that it is vulnerable to attack. The function that is used to drive Key stream is $\text{RC4}(v,k)$ as a function of the $v$ and the security key $k$. The plain text is a raw message supposed it is denoted by $M$, and after applying checksum $c$, it can be written as $P = (P \cdot c \cdot (M))$. On the other hand, the receivers, who have the same key, XOR the key stream with cipher text to get initial plain text [5][7]. The method to generate ciphertext shown in Figure No.1 is vulnerable to many attacks.
After seeing many drawbacks, IEEE introduced new method to generate cipher text in which Integrity Check (IC) field in the packet and 24 bit Initialization Vector (IV) field. In this method, for every packet, different RC4 key is used for example,

\[
\text{If } C_1 = P_1 \oplus \text{RC4}(v, k) \\
\text{and } C_2 = P_2 \oplus \text{RC4}(v, k) \\
\text{then } C_1 \oplus C_2 = (P_1 \oplus \text{RC4}(v, k)) \oplus (P_2 \oplus \text{RC4}(v, k)) = P_1 \oplus P_2.
\]

The above technique generates overhead on the computation while the security is still poor [5] [10].

WEP2, a stopgap enhancement to WEP, it uses 128 bit encryption and IV to generate ciphertext. Value of IV is large in WEP2 and this is drawback of WEP2 algorithm.

WEPplus is also known as WEP+, it provided better security by avoiding weak IVs drawbacks. To obtain complete effectiveness of WEP+, it is necessary to use it at both ends of wireless connections. Moreover, it is vendor specific. It is quite possible that strong threats like replay attacks can break it.

Wi-Fi Protected Access (WPA) was launched to remove security drawbacks in WEP. A new technique was adopted in this algorithm which is called Temporal Key Integrity Protocol (TKIP) which is used to rectify authentication and encryption loopholes in WEP [9]. It uses, Per packet key mixing function, Message Integrity Check (MIC), Initialization Vector with sequencing rules and Re-keying mechanism. WPA introduced one variation which is known as WPA Pre Shared Key, WPA (PSK). It provides strong encryption and encapsulation for authentication. If we combine both TKIP and WPA (PSK), then, the hacker will feel great trouble to find secret key.

Wi-Fi Protected Access 2 (WPA2) provides authentication, confidentiality and integrity to the wireless network. For better protection, it is based on Layer-2 of OSI model. It uses Counter-Mode with Cipher Block Chaining-MAC Protocol (CCMP). The CCMP uses Advance Encryption Standard (AES) encryption algorithm rather a weak RC4 encryption algorithm. For better protection against threats, it is necessary to use WPA2 for encryption and authentication.

### 3. WIRELESS NETWORK THREATS

As mentioned above, the wireless technology uses air as medium to communicate with each other. This medium made the wireless network susceptible to threats by attackers.

There are two types of attacks in wireless security:

- **Active Attacks**
- **Passive Attacks**

In Active Attacks, the attackers change the contents of the information and generate fake information in the network to destroy network security like Unauthorized Access, Active Eavesdropping, Man in the Middle Attack (MITM), Session Hijacking, Denial of Service (DoS), Replay, while in Passive Attacks, the attacker just listen to the traffic of the network, obtain information from the packets without changing it like passive Eavesdropping and Traffic Analysis. These types of attacks are very hard to detect.
3.1 Unauthorized Access

In Unauthorized Access, the user gains access to the network and can obtain data and use the bandwidth of the network easily. The attacker can violate the confidentiality and integrity of the network traffic by listening packets, changing in them as per requirements, send and receive the messages [1].

![Attacker](image)

**Fig. 2**

3.2 Denial of Service (DOS)

Denial of Service (DoS) is very famous attack to break down the wireless as well as wired network. To breakdown the network, it sends huge traffic on the Access Point make it unable to respond [6] [11].

In Wireless network, the DoS attack is just carried out by using a powerful enough transceiver, interference in shape of noise is generated to jam to the network.

Denial of Service (Dos) attack uses forged Disassociation technique to break network security. The attacker flood large number of disassociation frames to the client computer to break its connection with access point. After break-up of connection, the client computer again attempts to establish its connection with access point. It is pertinent to mention that client computer is already authenticated and only need association. To prevent re-association, the attacker continues to send Disassociation frames up to desired period.

Similarly, the attacker can use forged Deauthentication technique to breakup network security. The attacker flood Deauthentication frames to client computer and resulted, client computer unauthenticated with access point and will try to again authenticate itself with access point. To prevent again authentication, the attacker continue to send Deauthentication frames up to desired period.

Several tools are available over the internet like LANJack, Hunter_killer that can be used to launch DoS attack.

IEEE 802.11 wireless network standard by using the Medium Access Control (MAC) address, does not authenticate the source IP address, resultant, the attacker spoof MAC address and hijack the session. Furthermore, the Access Point does not prove itself as genuine Access Point.

3.3 Active Eavesdropping

The core threat of the wireless network is eavesdropping in which data in shape of signals is transferred from client workstation to access point.

In Active Eavesdropping, the attacker injects its data into the network signals to crack the secret key as shown in Figure No.3. The goal of this attack is to determine the
contents of message. The attacker can partially access the part of plaintext like source IP address, destination IP address etc or contents of the all traffic [1][6].

In Figure No. 3, the attacker takes advantages of WEP drawbacks where in WEP, the CRC is used to only check the integrity of the data not all the other contents of the packet like destination and source IP address. Hence, the attacker’s changes in these parameters cannot be detected [6].

IP Spoofing is example of Active Eavesdropping in which the attackers change destination IP address of the host that he controls.

3.4 Man in the Middle
The Man in the Middle is very dangerous attack wherein the attacker eavesdropping the communication and modifies it before sending [6]. Although, organization implemented VPN, SSH, IPSec security measures but these measures are breakable through MITM attack as these measures can only protect data confidentiality attack.

The attacker connects himself to the Access Point as a user and to a user as authentic Access point as shown in Figure No. 4. In this way, all the user data passed to the Access Point through the attacker and the attacker not only sniffs the data, but also can change the data, insert viruses in downloading files, change web pages setting easily.

IP Spoofing and Masquerading are the techniques that are used to make user fool. Moreover, encryption does not play security rule between Access Point and client user.
Address Resolution Protocol (ARP) poison is a technique that is used in Man in the Middle Attack. ARP enables the user to find users’ MAC address. When network receives a packet, whose destination IP address is missing or unknown, it cache the packet by sending this packet on network and ask to every client about it, if any machine match with it, then machine reply with ARP Reply packet and in this packet MAC address is available [1]. Once the cache has been updated, the attacker can act as MITM as shown in Figure No. 5.

![Fig. 5](image)

**3.5 Session Hijacking**

Session Hijacking indirectly resembles to the Man in the Middle Attack (MITM) wherein the attacker captures the session of the victim client [8] [11].

The victim just supposes that his/her session was expired due to whatever reasons while his/her session was handed over to attacker and he/she can exploit it as per his desire.

In Session Hijacking, the attacker, first obtains the MAC address of the victim and AP then it send MAC Disassociation message to the victim. The victim closes its session from network while his/her session in real opens in AP.

![Fig. 6](image)

The attackers, by using the victim MAC address, get the control of session. After getting control, the attacker use the said session for whatever purpose they desire. Session hijacking attack occurs in true but can continue up to extended time [1].
3.6 Replay
Replay attack also resembles to the Man in the Middle Attack with a little bit variation. In Replay attack is not real time attack as MITM attack, it grabs all the information (data and session) of the network and later on offline use the same to exploit the victim’s information [8].

The goal of attack is to access the network with the authorizations of the target without secret key as shown in Figure No.7 [1]. Moreover, the attacker may use resources of the network by using target client authorization and permission. This attack is real and often used to exploit the security of the wireless network.

3.7 Traffic Analysis
The Traffic Analysis is very simple and easy technique in which the attacker just notes the number and size of the packets transmitted over the air, protocols used by the network and active access point as shown in Figure No. 8 [1][8].

If packets are encrypted, then the attacker can get partial or full information from packet of the message. The attacker, before starting an active attack should get preliminary information about the network properties.

Wardriving is freeware tool that is available over the internet is very useful tool for traffic analysis. To find out target Access Point, the attacker moves from one place to another to find out active access points of the network [1]. Every Access Point (AP) broadcast its Service Set Identifier (SSID) over the air to recognize itself to the wireless notes desired to connect to network. Through this broadcasting information, access points allow everyone in its domain to identify them.
3.8 Passive Eavesdropping

This attack much resembles to traffic analysis attack. In this attack, the attacker gets the size of the packet, protocol used in wirelesses network, number of packets transmitted over air, and various characteristics of the packet as shown in Figure No. 9 [11].

In Passive eavesdropping attacks, the attacker exploits the victim’s privacy and information. In case of packet is encrypted then the attacker has to break its encryption to get plaintext or desired information from the packet [1].

WEP has various vulnerabilities because it has small numbers of initialization vectors (IV) sequences and this thing made it susceptible to attack. Rapidly reusing the same IV and made stream less vulnerable but can be cracked.

WPA2 use AES and 3DES strong encryption algorithms to ciphertex and in presence of these algorithms, passive eavesdropping is very hard. It can only be possible, if the packets are not encrypted.

4. THE HACKER’S TOOLBOX

To hack Wireless Network, several freeware tools are available on the internet. For better security in Wireless Network, the users must understand their methodologies and accordingly take security measures to protect their respective networks.

The table given below enlists some mostly commonly used freeware hacker’s tools with their description:
<table>
<thead>
<tr>
<th>TOOL</th>
<th>DESCRIPTION</th>
</tr>
</thead>
<tbody>
<tr>
<td>NetStumbler</td>
<td>NetStumbler is an active network scanner as well as freeware Wireless Access Point identifier that shows AP’s SSID, MAC address, type of encryption algorithm, Channels, noise, and signal strength. It is also used for fine-tuning a wireless link [7].</td>
</tr>
<tr>
<td>Kismet</td>
<td>Kismet is very advance level freeware wireless network diagnostic tool. It is used for passive eavesdropping attack. Kismet is used to monitor traffic, store data, sort data to identify SSID, channels, connection speed, MAC address, signal to noise ratio (SNR), graphically mapping of network by using GPS and range of IP addresses used in network. If hacker has multiple wireless network cards, then it spits out load to all of them to expedite the work [7].</td>
</tr>
<tr>
<td>AirSnort</td>
<td>AirSnort is widely used freeware tool that is used to break encryption of WEP. It is also used for passive monitoring of the wireless network as NetStumbler. After getting sufficient packets, it starts computing encryption to break the security [7].</td>
</tr>
<tr>
<td>Ethereal</td>
<td>It is freeware WLAN analyzer software that is used for passive eavesdropping. The interesting feature of said tool is its summary and detail summary for observed traffic.</td>
</tr>
<tr>
<td>WEPcrack</td>
<td>This tools name depicts its function WEPcrack is wireless network cracking tool that used to exploit the vulnerabilities in RC4 algorithm by using latest discovered drawbacks in 802.11.</td>
</tr>
<tr>
<td>THU-RUT</td>
<td>Freeware WLAN discovery tool. It used <em>brute force</em> attack to obtain low traffic access points. *&quot;Your first knife on a foreign network.&quot;*</td>
</tr>
<tr>
<td>WEPWedgie</td>
<td>It is used for active eavesdropping attack as it used to determine WEP Keystreams and inject with known Keystreams. It has pingscanning and portscanning facility via injection channels.</td>
</tr>
<tr>
<td>AirSnarf</td>
<td>It is AP spoofing tool that is used to deceive users by forwarding their important/sensitive information to attacker. To victimize user, AirSnarf imitate legitimate AP and to create the same login page that normally be displayed by the access point. When the user logsins, its login, information will be sent to attacker and he/she will use it as per his/her desire. This is very simple but effective tool for wireless network hacking.</td>
</tr>
<tr>
<td>Airjack</td>
<td>Airjack is very dangerous tool for wireless network. It is used for Denial of Service (DoS) and Man in the Middle (MITM) attacks as it begins with sending disassociate or de-authenticates frames at very high speed to access point and accordingly AP drops connections with its all connected users.</td>
</tr>
</tbody>
</table>
### TOOL | DESCRIPTION
---|---
Cain&Abel | Password recovery tool that is used to recover passwords by cracking encryption. It uses Dictionary attack, Brute-Force, and Cryptanalysis attacks. It is also very effective and easy to operate freeware tool to hack wireless network within few minutes.
HostAP | It is firmware for Prism cards to act as AP. HostAP has multiple scanning, broadcasting and managements options that is used to connect disconnected clients with the HostAP and after that the attacker can grab all the information as per his/her requirements.
Hotspotter | Passively monitor all the activities of the wireless network.
ASLEAP | It is toolkit used to recover LEAP networks passwords. It is also used to de-authenticate users from network and force them to connect with rogue AP.
WEP Attack | WEP Attack tool is used for Brute-Force. WEP cracking by using Dictionary attacks against WEP keys.
AiroPeek | Packet Analyzer specially IEEE 802.11b. It is used for evaluate network performance, signal strength, number of channels, and speed/data rate.

## 5. COUNTERMEASURES

Although, there are various flaws in WEP algorithm but still it is possible for users to secure their respective wireless networks. To deal with above mentioned security threats, the following techniques are recommended to minimize the security risks involved in wireless network.

### 5.1 Training and Educating Users

The first step in wireless network security is to educate the users about how to secure network. It is often observed that end users does not know how to implement security and leaves various loopholes for attackers. If users will well aware about wireless tools configurations/settings and how to secure their respective network, then it is quite possible to reduce security risks.

### 5.2 Wireless Network Auditing

It is powerful technique to secure wireless network. The user should scan his/her network through network scanner to know about the activities of network [6]. Several free network scanning softwares like NetStumbler and Kismet are available over the internet.

### 5.3 Turn Off AP When You Would Not Use It

If the user turns off his/her wireless network router/access point, when he/she is not using it then we can limit the time that it is susceptible to hack [6].
5.4 Change Router’s Pre-Set Password

Every manufacture of the wireless router/access point set default user name and password. If the user does not change it, then it is very sweet cake for the attacker as the attacker simply scans access point and accesses it through its default username and password.

Therefore, it is strongly recommended that in first instance, the user should to change by default username and password.

5.5 Change SSID

Every access point has by default ID and attacker can easily find access point by entering default ID. All the devices that connect to wireless network use same SSID. If user does not change default SSID, then it is like to leave default password. Moreover, it is also best practice to change SSID within 30 days or before.

5.6 Turn Off SSID Broadcasting

Access Point uses SSID broadcasting technique to show its presence in the environment in which it is working. By turning off SSID broadcasting function, it is very hard for the attacker to scan the network [6]. Therefore, it is recommended to turn off this function if router allows this.

5.7 Utilize Virtual Private Network

In securing wireless network, it is very best security technique to put behind wireless access point to Virtual Private Network (VPN). A Virtual Private Network is very good solution to authenticate unauthorized users who try to connect themselves to the network and encrypt their respective communication.

If an organization has more than one access points, then it is strongly recommended that connect all of them with one common switch and connect this switch to VPN server.

5.8 Location of Access Point

During designing of wireless network, it is suggested to always place access point in the middle of the network place to avoid its unnecessary signal broadcasting outside the network place. For better security, the signal strength and power level should to set as per boundary of the network.

5.9 Use of Antivirus, Firewall and Anti Spyware

To exploit the security, viruses and Malwares play their roles accordingly. For better security, it is recommended that the client must install updated firewall, antivirus, anti-malware and Anti Spyware softwares on his computers.

5.10 Encryption

The best way to secure wireless network from unauthorized users/attackers is to use encryption for ciphertext. WEP algorithm has various vulnerabilities and it is not recommended to use it for encryption purpose. IEEE 802.11i use Advance Encryption Standard (AES) and Triple Data Encryption standard (3DES).
5.11 SSH
Most users use telnet utility for remote connection which is insecure in wireless network, the users should use SSH with tunneling feature to provide secure remote connection.

5.12 Third Party Wireless Security Tools
Various third party security tools are available over the internet to provide better security for wireless networks. Some of them are given below:

5.12.1 Nextcomm, Inc
It uses MD5 hash algorithm to generate Keystream. NextComm provides this facility through IC chip which should be part of access point and wireless card. MD5 has rapidly changed Keystream which is also known as Key hopping to prevent attackers to get enough information about analysis.

5.12.2 Interlink Networks
It provides wireless security through RADIUS for secure authentication.

5.12.3 Airdefense
It is very powerful and famous wireless security tool which provides intrusion protection, show vulnerabilities, prevent network from attackers and also provides assistance to users, improve their security and performance.

5.13 Enable MAC Filtering
MAC filtering is very famous and best technique to only authenticate clients which are part of the network.

5.14 Upgradation of Access Point Firmware
The older Access Points (APs) have no capability to support latest security algorithms. The client should to update its Access Point (AP) firmware software to WPA2.

5.15 Turn Off Dhcp Server
The attacker while connecting attacking required IP address of the network. If DHCP server will enable then it is sweet cake for attacker. Therefore, for better security measures, it is suggested to assign static IP address to every client computer and turn off DHCP server on Access Point.

6. COMMENTS AND CONCLUSION
Wireless technology is very famous all over the world due to its low cast and easy to install characteristics, although it has numerous security flaws but still it captures the market.

This paper discusses brief taxonomy about the Wireless Equivalent Privacy (WEP) vulnerabilities, wireless security threats and attacks, hacking tools which are freeware and available on the internet to hack wireless network and also describes various
countermeasures techniques to secure wireless network. By understanding these attacks techniques with freeware tools and its countermeasures make a user to understand about the said risks and how to mitigate them. I have indicated several existing freeware hacking tools that implement attack techniques to exploit the weaknesses in the protocol designs. Moreover, I have pointed out several best practices that can mitigate the insecurities.

REFERENCES

REIFYING TECHNO-E LEARNING AS CATALYST FOR AFFIRMATIVE CHANGE

Syed Faizan Haider¹, Daniyal Alghazawi² and Waqas Ahmad³

¹ IS Department, King Abdulaziz University, Jeddah Saudi Arabia / Superior University, Lahore, Pakistan. Email: hahohay63@yahoo.com
² IS Department, King Abdulaziz University, Jeddah Saudi Arabia
³ Department of Quality assurance, Institute of Space Technology, Islamabad, Pakistan

ABSTRACT

Man, the vicegerent of Almighty (The creator of Heavens & Earth & all beings in these) has been endowed with the title of supremacy over the angels by blessing him with the knowledge. The knowledge of perceiving things, for their trough and crust behavior, usability and application etc.

There has been various modes of attaining education Verbal and Written for example, since the times immemorial to the currently prevailing like:

- E Learning
- Print Media
- Practical & Demonstration
- Illustration & Multimedia.

Amongst the above mentioned, e-learning and multimedia is the strong force through which vast knowledge can be conveyed to the learner in a relatively short time thus can be attributed as a positive change agent.

This paper is an attempt to explore and present aforementioned idea and to some extent implement it through the use of interactive and e-learning and thereby enhancing student participation that would eventually lead to:

- Interest Maintaining
- Thought Improvement
- Creativity Enhancement

especially with reference to the Pakistani educational environment.

KEYWORDS

Supremacy; Perception; Usability; Creativity; Educational Environment; Multimedia.

1. INTRODUCTION

Human being basically blessed with five senses through the five organs namely Ear, Eyes, Skin, Tongue, and Nose. These divine endowed senses have led the human being to set his journey from perception [3] to awareness then forming hypothesis and later on heading towards education that formed its rout to knowledge and finally to the learning level. The connectivity vs. understanding is as follow
The ever-inquisitive nature of the reformers, thinkers, philosophers and scientists made learning to evolve in various forms i.e. M-Learning [12], learning by heart (Verbal Learning), Learning through writing on stones, Leaves & Leather etc. to the printing and later on to Vocal (Radio), Vision (TV) and different other styles of learning up to the current form termed as new media.

2. DEFINITION OF MULTIMEDIA:

Different definitions [4], [7], [9] etc. have evolved over the previous several years, in my opinion it could be defined as:

The integration of diversified; Print, Newspaper, Radio, Communication, TV, and Web Media if converged leads towards a High-Tech media (Multimedia).
Media is a strong force through which very huge amount of knowledge can be conveyed to pupils in a very short time. It motivates student interaction, experimentation, and cooperative learning [1].

e-Learning and Multimedia could play an effective role in enhancing various intellectual capabilities like:

![Multimedia Diagram](image)

**Fig. 3**

The changes brought through multimedia in the learning environment are not only affirmative but also more evolutionary than revolutionary.

### 3. TECHNOLOGY & METHODOLOGY

The research under consideration consists of two parts. First one deals with finding out the basic needs and drawback of utilization of Multimedia at various levels of education and to suggest some remedies for this problem. The second part consist of the developing partial English language learning software that would be a source of self-learning dealing with Pronunciation, Punctuation, Grammar, firstly from Nursery & Primary level and then from Primary to higher classes and so on so forth.

Hardware and Software requirement include the software itself along with a computer with optimal 800 MHz speed, with proper Audio and video card, speakers, CD Rom Drive, High resolution colored Monitor / multimedia projector.

### 4. RESULT & DISCUSSION

If we have to consider multimedia to be an inevitable tool in every aspect of life in general and in the dissemination and enhancing of knowledge and awareness in particular, then multimedia consideration could be for several reasons for its capabilities as shown in figure below.
Furthermore as a tool that help to think deep leading to pursuing curiosity and intelligence, enhancing in the students to visualize their ideas on any topic [8], [10]. Multimedia is seen as one way of solving the growing problems between the skills and availability of the work force and the increasing demand of technology in industrialized countries. [11]

New media can bring a drastic change in the study habits of students when they use the computer through keyboard. It can help the trainees to

- Acquire information and skills.
- Practice these skills.
- Reinforce their skills in the course of their activities

Pakistan is a developing country with 23.10 % of people (Adults) living below poverty line which has been drawn for the year 2004, as Rs. 848.80 per adult equivalent per month [6] which is equivalent to 14.14 $, ridiculous isn’t it? ($1.25 @Day as of 2011) and this makes more than 50% of Pakistani population under the poverty line [6 a]
Now this poverty level has been associated with economic growth, which implies greater availability of public resources towards the improvement of both quantity and quality of education and health services at the macro level.

As is evident from fig-5 that number of professionals (P/colleges and University) in public sector in Pakistan is 382 and 29 respectively [5] with the enrolment in same category being 1.63 and 1.27 lakhs, that raised to 420 and 61 in number with the enrolment in same category being 2.96 and 2.07 lakhs [5 a] total inclusive of both male & females shown in fig-6, where as the number of the universities/degree awarding institutions in the private sector is 30, with their enrollment total male & female was 16410 in the year 2003-04 [6] and that got up to the level of 56 enrolling 87,310 in the year 2006-2007 [6 a]
Now this above-mentioned total enrollment was in all the disciplines namely Agriculture, Engineering, Medical, Law and Education. This all other category comprises of Tibb, Homeopathy, Fine arts and Computer, as shown in fig-7 below.

![Professional Enrolment Category wise](image)

Others includes Tibb, Homeopathic, Fine Arts and Computer

**Fig. 7**

If we just assume the computer science category where multimedia could be applied, to be 25% of the other disciplines that turns out to be around 11,000, which is in the recession in the current year with only less than 500 teachers in all the branches on the computer science serving all over the country. The situation has not improved till 2006-7 that was less than 12,600 & the teachers 710 teachers which is relatively a healthy sign and improving further.

The question arises that if this number is considered to be only the experts in multimedia education; does this number serve the purpose of teaching, if multimedia style of education is to be proposed at the primary level?

The discipline e-learning & multimedia (IT and Computer Science) is open ended; anybody could enter into it even if he/she has taken only one course or even without a course by having a book of computer applications and its utilities that has led to the artificial recession in the field. The need is that only the professionals with proper qualification and expertise to be put in place to work for the progress and enhancement in the said field.

Currently only a few institutions in the public and private sector are trying to attract the attention of richer class to send their offspring to such heavily tuitioned institution to get into this unrealized stream which eventually in the years to come would be likely to convert into an unfathomable ocean.
5. SUGGESTION & CONCLUSION

The advancement of technology has no bounds both in terms of its natural up and down trends in the days and the years to come. E-Learning and Multimedia has led towards interdisciplinary collaborative Hybridism. People retain only 20% of what they see and 30% of what they hear but they remember 50% of what they see, hear and as much as 80% what they see and hear and do simultaneously. That’s why multimedia provides such a powerful tool for teaching, training and learning. [4]

<table>
<thead>
<tr>
<th>Perception Contribution (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sight</td>
</tr>
<tr>
<td>Hearing</td>
</tr>
<tr>
<td>Touch</td>
</tr>
<tr>
<td>Smell</td>
</tr>
<tr>
<td>Taste</td>
</tr>
</tbody>
</table>

The words of the Lord of the worlds: “Who teacheth by the pen, Teacheth man that which he knew not” [13]

It is above board that what so ever the sources of attaining and getting knowledge be there but still the knowledge/education sought can’t be authentic without the use of pen i.e. the physical involvement of writing with other perception sensors is as important as heart in the body.

The idea of introduction of multimedia in the Pakistani educational environment sounds to be good one in the cities like Lahore, Karachi, Islamabad, Peshawar, Quetta to name a few, for the elite or rich class (Both male & Female) [2] studying in the private sector institutions. We need first of all to homogenize and equalize the double or rather many fold standards towards education but for this purpose we have to go for:

- The eradication of Poverty
- Reforming the Political and Social System
- Elevating the Literacy level
- Alleviating the Vedera System
- Excluding red tapeism.

REFERENCES

WEIGHTED ESTIMATORS OF POPULATION MEAN USING TWO AUXILIARY VARIABLES UNDER TWO PHASE SAMPLING

Aamir Sanaullah and Muhammad Hanif
National College of Business Administration, Lahore, Pakistan
Email: chaamirsanaullah@yahoo.com; hanif@lums.edu.pk

ABSTRACT

To estimate the population mean $\bar{Y}$ of the variate $y$ under study using two phase sampling procedure, a class of new estimators has been proposed when the population means of main auxiliary variable $X$ and secondary auxiliary variable $Z$ is available. The class of new estimators has less mean square of error than the mean squares of some previous class of estimators. The theoretical results obtained have been illustrated by taking some empirical study.

KEY WORDS

Two phase sampling; Mean-Squared Error; Bias; Auxiliary Variable; Simple random Sampling without Replacement.

1. INTRODUCTION

In sampling theory an unremitting concern is to estimate population mean by ratio, product and regression methods of estimation in the presence of multi-auxiliary variables using two phase sampling procedure. In this context many of the ratio, product and regression type estimators has been proposed from time to time for single and double sampling. Double sampling was first used by Neyman (1938) to obtain information on strata size in stratified sampling. Sukhatme (1962), Unnikrishan & Kunte (1995) made several efforts to provide a lot of applications of two phase sampling.

When regression line passes through the origin and correlation between study variable $Y$ and auxiliary variable $X$ is positive, the ratio method of estimation is regarded as the most practicable for estimating the population mean, however in the case of negative correlation, the product method of estimation is employed quite effectively. By considering the work of Searls (1964), Das and Tripathi, (1978), Sen (1978), Das and Tripathi, (1980), Sisodia and Dwivedi (1981), Panday and Dubey (1988), Upadhyaya and Singh (1999), Khoshnevisan et al. (2007) and Sing et al. (2007), proposed ratio type estimators for population mean by using different known information of certain parameters e.g. coefficient of correlation, coefficient of variation and kurtosis with known mean of the auxiliary variables to improve the efficiency of the estimators. The product estimator is developed by Robson (1957) and it is rediscovered by Murthy (1964). Rao and Mudholkar (1967) and Srivastava (1971) and others have made the extension of the ratio estimators to the case where multiple auxiliary variables are used to increase the precision. An improved estimator was proposed by Hanif et al. (2010) using multi auxiliary variables to estimate mean of population in multiphase sampling and the
estimator was compared with other estimators in which full information was available in double sampling. The proposed estimator was efficient than Abu-Dayyeh and other estimators.

1.1 Expectation of Errors in Two Phase Sampling Procedure

Let a population of $N$ units say $U = \{U_1, U_2, \ldots, U_N\}$. Let $y$ and $(x, z)$ be the variate of interest and auxiliary characteristics respectively related to $y$ assume real non-negative $j$th value $(y_j, x_j, z_j) j = 1, 2, \ldots, N$ with population means $\bar{Y}, \bar{X}, \text{ and } \bar{Z}$ respectively. A large first phase sample of size $n_1$ units is selected from $N$ units from the population by Simple random sampling without replacement (SRSWOR) and the characteristics $x$ and $z$ say $(x_1, z_1)$ are measured on it. A smaller second phase sample of size $n_2$ is selected from $n_1$ by SRSWOR and the characters $y$ and $z$ say $(y_2, z_2)$ are measured on it.

For a simple random sampling without replacement (SRSWOR), we have some assumptions as following,

$$
E(e_{(y_1)}) = E(e_{(z_1)}) = E(e_{(x_1)}) = 0
$$

$$
E(e_{(y_1)\|^2}) = \theta_2 \bar{Y}^2 C^2_y \\
E(e_{(z_1)\|^2}) = \theta_2 \bar{Z}^2 C^2_z
$$

(1.1.1)

$$
E(e_{(x_1)\|^2}) = \theta_1 \bar{X}^2 C^2_x \\
E(e_{(z_1)\|^2}) = \theta_1 \bar{X}^2 C^2_y \\
E(e_{(y_1)\|^2}) = \theta_1 \bar{X}^2 C^2_x \\
E(e_{(z_1)\|^2}) = \theta_1 \bar{X}^2 C^2_y
$$

2. SOME AVAILABLE ESTIMATORS

We have presented some of the well known estimators for population mean which use the information on auxiliary information. Mean squares of errors have also been given with each of the estimators.

Samiuddin and Hanif’s (2006) Chain Ratio Estimator-I

$$
t_1 = \frac{\bar{y}_2}{\bar{X}} \frac{\bar{z}_1}{\bar{Z}}
$$

(2.1)

$$
MSE(t_1) = \bar{Y}^2 \left[ \theta_1 \left( C_x^2 + 2C_x (C_y \rho_{xy} + C_z \rho_{xz}) \right) + \theta_2 \left( C_y^2 + C_z \left( C_z + 2C_y \rho_{yz} \right) \right) \right]
$$

(2.2)

Samiuddin and Hanif’s (2006) Chain Ratio Estimator-II

$$
t_2 = \frac{\bar{y}_2}{\bar{X}} \frac{\bar{z}_2}{\bar{X}_1}
$$

(2.3)

$$
MSE(t_2) = \bar{Y}^2 \left[ \theta_2 \left( C_y^2 + C_z^2 - 2C_y C_z \rho_{yz} \right) + \theta_1 \left( C_x^2 - 2C_y C_x \rho_{xy} + 2C_z C_x \rho_{xz} \right) \right]
$$

(2.4)
Khaire and Srivastava (1981) Regression cum Ratio Estimator

\[ t_3 = \left[ \frac{\bar{y}_2 + b_{yx} (\bar{X} - \bar{x}_1)}{z_2} \right] \frac{Z}{\bar{x}_2} \]  

\[ MSE(t_3) = \bar{Y}^2 \left[ \theta_2 \left( C_x^2 + C_z^2 - 2C_xC_z\rho_{yz} \right) - \theta_1 \theta_3 \rho_{xy} \left( C_y \rho_{xy} - 2C_z \rho_{xz} \right) \right] \]  

Singh et al. (2005) Ratio cum Product Estimator

\[ t_4 = \bar{y}_2 \left( \frac{\bar{X} + \rho_{xz}}{x_1 + \rho_{xz}} \right) \left( \frac{\bar{z}_1 + \rho_{xz}}{Z + \rho_{xz}} \right) \]  

\[ MSE(t_4) = \theta_1 \bar{Y}^2 \left[ C_x^2 + C_z^2 \mu_1^2 + C_z^2 \mu_2^2 - 2\mu_1 C_x C_z \rho_{xy} - 2\mu_2 C_y C_z \rho_{yz} + 2\mu_1 \mu_2 C_x C_z \rho_{xz} \right] \]  

3. 1st THE SUGGESTED ESTIMATOR

Motivated by Samiuddin and Hanif’s (2006) chain ratio estimators, using x and known auxiliary variables, consider \( t_1 = \frac{\bar{y}_2}{x_1} \frac{\bar{x}_1}{\bar{x}} \) and \( t_2 = \frac{\bar{y}_2}{x_1} \frac{\bar{z}_1}{\bar{z}} \) as

\[ \hat{Y}_1 = \delta t_1 + (1-\delta) t_2 \] is a new proposed estimator with a constraint \( 0 < \delta < 1 \) as

\[ \hat{Y}_1 = \delta \frac{\bar{y}_2}{x_1} \frac{\bar{x}_1}{\bar{x}} + (1-\delta) \frac{\bar{y}_2}{x_1} \frac{\bar{z}_1}{\bar{z}} ; \quad 0 < \delta < 1 \]  

where \( x_1 \) and \( z_1 \) are the sample means from sample at 1st and \( x_2, z_2 \) and \( y_2 \) are the means from sample at 2nd phase. \( \delta \) and \( 1-\delta \) are weights attached with \( t_1 \) and \( t_2 \) respectively.

To obtain Bias and MSE,s, consider

\[ \bar{x}_1 = \bar{X} + e_{x_1} ; \quad \bar{y}_2 = \bar{Y} + e_{y_2} ; \quad \bar{z}_1 = \bar{Z} + e_{z_1} ; \quad \bar{z}_2 = \bar{Z} + e_{z_2} \]

Thus (3.1) is the expended form to obtain MSE,s to the first degree of approximation.

\[ \hat{Y}_1 = \bar{Y} + e_{y_2} + \delta \frac{\bar{Y}}{Z} e_{z_1} + (2\delta - 1) \frac{\bar{Y}}{X} e_{x_1} + (\delta - 1) \frac{\bar{Y}}{Z} e_{z_2} \]  

Now MSE can defined as

\[ MSE(\hat{Y}_1) = E(\hat{Y}_1 - \bar{Y})^2 \]

\[ = [e_{y_2} + \delta \frac{\bar{Y}}{Z} e_{z_1} + (2\delta - 1) \frac{\bar{Y}}{X} e_{x_1} + (\delta - 1) \frac{\bar{Y}}{Z} e_{z_2}]^2 ; \]

\[ = \bar{Y}^2 [\theta_2 C_x^2 + (2\delta^2 - \delta) \theta_1 C_x^2 + (2\delta - 1)^2 \theta_2 C_x^2 + (\delta - 1)^2 \theta_2 C_x^2 + \theta_1 \theta_3 \rho_{xy} C_x C_y + (2\delta - 1) \theta_1 \rho_{xy} C_x C_y + (2\delta - 1)^2 \theta_1 \rho_{xy} C_x C_y] \]  

(3.3)
where
\[ \theta_1 = \frac{1}{n_1} - \frac{1}{N}; \quad \theta_2 = \frac{1}{n_2} - \frac{1}{N} \]
and
\[ \delta = \frac{\theta_1 A_1 + \theta_2 A_2}{\theta_1 A_3 + \theta_2 A_4}; \quad A_1 = C_x^2 + 4C_x^2 - 2\rho_{xy}C_xC_y + 3\rho_{xy}C_yC_z; \quad A_2 = 2C_z^2 - \rho_{yx}C_xC_z \]
\[ A_3 = 4C_z^2 + 8C_z^2 + 8\rho_{yx}C_xC_z; \quad A_4 = 2C_z^2 \]

Now Bias can be obtained by using 2\textsuperscript{nd} degree approximation as
\[
\text{Bias}(\hat{Y}_1) = E(\hat{Y}_1) - \bar{Y} = \bar{Y} \left[ \theta_2 C_y^2 + (1-\delta)\theta_1 C_x^2 - \delta \theta_2 C_x^2 + \theta_1 \rho_{xy} C_xC_y \right. \\
\left. - \theta_2 \rho_{yx} C_yC_z + (2\delta - 1)\theta_1 \rho_{xy} C_xC_y \right]
\]
(3.4)

**Remark 3.1**

If \( \delta \to 0 \), the new suggested estimator \( \hat{Y}_1 \) will have very close resemblance of Samiuuddin and Hanif (2006) estimator \( t_2 \), i.e.
\[
\hat{Y}_1 = \delta t_1 + (1-\delta) t_2 \to \quad t_2 = \frac{\bar{X}}{\bar{x}_1} \frac{\bar{Z}}{\bar{z}_2}
\]
The empirical relation \( \text{MSE}(\hat{Y}_1) < \text{MSE}(t_2) \) do holds good, so for the reason \( \hat{Y}_1 \) is preferred over \( t_2 \).

**Remark 3.2**

If \( \delta \to 1 \), the new suggested estimator \( \hat{Y}_1 \) will tend to approximate Samiuuddin and Hanif (2006) estimator \( t_1 \), i.e.
\[
\hat{Y}_1 = \delta t_1 + (1-\delta) t_2 \to \quad t_1 = \frac{\bar{y}_2}{\bar{x}_1} \frac{\bar{z}_1}{\bar{z}_2}
\]
The empirical relation \( \text{MSE}(\hat{Y}_1) < \text{MSE}(t_1) \) do holds good, so for the reason \( \hat{Y}_1 \) is preferred over \( t_1 \).

### 4. 2\textsuperscript{nd} SUGGESTED ESTIMATOR

Using \( t_1 = \frac{\bar{y}_2}{\bar{x}_1} \frac{\bar{z}_1}{\bar{Z}} \) and \( t_2 = \frac{\bar{X}}{\bar{x}_1} \frac{\bar{Z}}{\bar{z}_2} \) for known auxiliary information, another new estimator is proposed as \( \hat{Y}_2 = \delta t_2 + (1-\delta)t_1 \) with a constraint \( 0 < \delta < 1 \). The new estimator can be written as
\[
\hat{Y}_2 = \delta \frac{\bar{y}_2}{\bar{x}_1} \frac{\bar{z}_1}{\bar{z}_2} + (1-\delta) \frac{\bar{y}_2}{\bar{x}_1} \frac{\bar{z}_1}{\bar{z}_2}; \quad 0 < \delta < 1
\]
(4.1)
Thus (4.1) is the expended form to obtain MSE,s to the first degree of approximation.

$$\hat{Y}_2 = \bar{Y} + e_{\hat{Y}_2} + (1 - \delta) \frac{\bar{Y}}{\bar{Z}} e_{\bar{z}_1} + (1 - 2\delta) \frac{\bar{Y}}{\bar{X}} e_{\bar{x}_1} - \delta \frac{\bar{Y}}{\bar{Z}} e_{\bar{z}_2} \quad (4.2)$$

Now MSE can be defined as

$$MSE(\hat{Y}_2) = E(\hat{Y}_2 - \bar{Y})^2$$

$$= \left[ e_{\hat{Y}_2} + (1 - \delta) \frac{\bar{Y}}{\bar{Z}} e_{\bar{z}_1} + (1 - 2\delta) \frac{\bar{Y}}{\bar{X}} e_{\bar{x}_1} - \delta \frac{\bar{Y}}{\bar{Z}} e_{\bar{z}_2} \right]^2$$

$$= \bar{Y}^2 \left[ (1 - 3\delta + 2\delta^2) \theta_1 e_{\bar{z}_1}^2 + (1 - 2\delta) \theta_2 e_{\bar{x}_1}^2 + (1 - 2\delta) \theta_3 e_{\bar{z}_2}^2 + \theta_2 C_x C_z + (2\delta - 1)^2 \theta_1 \rho_{xz} C_x C_z \right.$$ \n
$$\left. + (1 - \delta) \theta_1 \rho_{yz} C_y C_z + (1 - 2\delta) \theta_1 \rho_{xy} C_x C_y - \delta \theta_2 \rho_{yz} C_y C_z \right] \quad (4.3)$$

$$\delta = \frac{\theta_1 B_1 + \theta_2 B_2}{\theta_1 B_3 + \theta_2 B_4};$$

where

$$B_1 = C_x^2 + 4C_z^2 - 4 \rho_{yz} C_y C_z + 3 \rho_{zy} C_y C_z; \quad B_2 = 2C_z^2 - \rho_{yz} C_y C_z$$

$$B_3 = 4C_z^2 + 8C_y^2 + 8 \rho_{zy} C_y C_z; \quad B_4 = 2C_z^2$$

Now Bias can be obtained by using 2nd degree approximation as

$$Bias(\hat{Y}_2) = E(\hat{Y}_2) - \bar{Y} = \bar{Y} \left[ \delta (\theta_2 C_z^2 + \theta_1 C_x^2) + \theta_1 \rho_{xz} C_x C_z \right.$$ \n
$$\left. + (\theta_1 - 2\theta_2) \theta_1 \rho_{xy} C_x C_y + (\theta_1 - \delta \theta_1 - \theta_2) C_y C_x \rho_{yz} \right] \quad (4.4)$$

**Remark 4.1**

If $\delta \to 0$, the new suggested estimator $\hat{Y}_2$ will have very close resemblance of Samiuddin and Hanif (2006) estimator $t_2$, i.e.

$$\hat{Y}_2 = \delta t_2 + (1 - \delta) t_1 \quad \rightarrow \quad t_1 = \frac{\bar{y}_2}{\bar{X}} \frac{\bar{z}_1}{\bar{Z}}$$

MSE$$(\hat{Y}_2) < $MSE$(t_1)$ and the empirical relation to hold s good, so for the reason $\hat{Y}_2$ is preferred over $t_1$.

**Remark 4.2**

If $\delta \to 1$, the new suggested estimator $\hat{Y}_2$ will tend to approximate Samiuddin and Hanif (2006) estimator $t_2$, i.e.

$$\hat{Y}_2 = \delta t_2 + (1 - \delta) t_1 \quad \rightarrow \quad t_2 = \frac{\bar{y}_2}{\bar{X}} \frac{\bar{Z}}{\bar{z}_2}$$

MSE$$(\hat{Y}_2) < $MSE$(t_2)$ and the empirical relation do holds good, so for the reason $\hat{Y}_2$ is preferred over $t_2$. 
6. EMPIRICAL STUDY

To comprise a forceful inspiration about the gain in efficiency of the proposed family over estimators in the literature, we take the four observed populations. The source of population, variable y, auxiliary variables X and Z, population size N, sample size at each phase (n1, n2) are given in table A-1 in appendix. The comparison is made with the estimators of population mean by Samiuddin and Hanif (2006), Khair and Srivastava (1981), and Sing et al. (2005). Mean squares of errors for these estimators and proposed estimators are given in table A-2 in appendix. The relative percent efficiency of the proposed estimators over other well known estimators is given in table-1 given below.

Table 1:
Percentage Relative Efficiencies of proposed estimators $\hat{Y}_1$ and $\hat{Y}_2$ with $t_1$, $t_2$, $t_3$ and $t_4$

<table>
<thead>
<tr>
<th>Estimator</th>
<th>Population-I</th>
<th></th>
<th>Population-II</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\hat{Y}_1$</td>
<td>$\hat{Y}_2$</td>
<td>$\hat{Y}_1$</td>
<td>$\hat{Y}_2$</td>
</tr>
<tr>
<td>$t_1$</td>
<td>784.602</td>
<td>819.1344</td>
<td>775.0546</td>
<td>761.1861</td>
</tr>
<tr>
<td>$t_2$</td>
<td>442.992</td>
<td>462.4889</td>
<td>320.9655</td>
<td>315.2222</td>
</tr>
<tr>
<td>$t_3$</td>
<td>410.005</td>
<td>428.0502</td>
<td>370.8699</td>
<td>364.2337</td>
</tr>
<tr>
<td>$t_4$</td>
<td>442.081</td>
<td>461.5381</td>
<td>319.3601</td>
<td>313.6455</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Estimator</th>
<th>Population-III</th>
<th></th>
<th>Population-IV</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\hat{Y}_1$</td>
<td>$\hat{Y}_2$</td>
<td>$\hat{Y}_1$</td>
<td>$\hat{Y}_2$</td>
</tr>
<tr>
<td>$t_1$</td>
<td>756.265</td>
<td>781.9358</td>
<td>633.808</td>
<td>630.3374</td>
</tr>
<tr>
<td>$t_2$</td>
<td>398.4089</td>
<td>411.9326</td>
<td>349.0461</td>
<td>347.1348</td>
</tr>
<tr>
<td>$t_3$</td>
<td>398.813</td>
<td>412.3512</td>
<td>313.5134</td>
<td>311.7967</td>
</tr>
<tr>
<td>$t_4$</td>
<td>397.3514</td>
<td>410.8391</td>
<td>347.6414</td>
<td>345.7378</td>
</tr>
</tbody>
</table>

7. CONCLUSION

The results of comparative study in the tables 1 are clearly indicating that the suggested estimators are more efficient than Samiuddin and Hanif (2006), Khair and Srivastava (1981), and Sing et al. (2005).

REFERENCE

Table A-1
Description of Populations

<table>
<thead>
<tr>
<th>S#</th>
<th>Population</th>
<th>Study Variable (Y)</th>
<th>Main Auxiliary Variable (X)</th>
<th>Secondary Auxiliary Variable (Z)</th>
<th>(N, n₁, n₂)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Population Census Reports of Okara District (1998) Pakistan</td>
<td>Population Matric and above</td>
<td>Primary and Below</td>
<td>Population both Sex</td>
<td>(300, 60, 12)</td>
</tr>
<tr>
<td>3</td>
<td>Advanced Applied Linear Model (Nachstshemim, Neter and Kutner, 2004)</td>
<td>Total Cost of Claims</td>
<td>Total no of Interventions</td>
<td>No of Emergency Room Visits</td>
<td>(778, 156, 31)</td>
</tr>
</tbody>
</table>

Table A-2
Mean Square Errors (MSE) of Estimators

<table>
<thead>
<tr>
<th>Population</th>
<th>$\hat{Y}_1$</th>
<th>$\hat{Y}_2$</th>
<th>$t_1$</th>
<th>$t_2$</th>
<th>$t_3$</th>
<th>$t_4$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>31.46437381</td>
<td>30.13794687</td>
<td>246.8702791</td>
<td>139.3846517</td>
<td>129.0055466</td>
<td>139.0981172</td>
</tr>
<tr>
<td>2</td>
<td>17.12605656</td>
<td>17.43808743</td>
<td>132.7362899</td>
<td>54.96872815</td>
<td>63.51538271</td>
<td>54.69378449</td>
</tr>
<tr>
<td>3</td>
<td>23.4364804</td>
<td>22.66706792</td>
<td>177.2419272</td>
<td>93.37303344</td>
<td>93.46793011</td>
<td>93.12517604</td>
</tr>
<tr>
<td>4</td>
<td>9.727675165</td>
<td>9.781235049</td>
<td>61.65478076</td>
<td>33.95407282</td>
<td>30.49756975</td>
<td>33.81743084</td>
</tr>
</tbody>
</table>