

IDENTIFICATION OF PERIODIC AUTOREGRESSIVE MOVING-AVERAGE TIME SERIES MODELS WITH R

¹Hazem I. El Shekh Ahmed, ²Raid B. Salha and ³Diab I. AL-Awar

¹Department of Mathematics, Quds Open University, Gaza, Palestine

²Department of Mathematics, Islamic University, Gaza, Palestine

³Department of Mathematics, Al-Aqsa University, Gaza, Palestine

Received 2014-03-01; Revised 2014-03-04; Accepted 2014-08-08

ABSTRACT

Periodic autoregressive moving average PARMA process extend the classical autoregressive moving average ARMA process by allowing the parameters to vary with seasons. Model identification is the identification of a possible model based on an available realization, i.e., determining the type of the model with appropriate orders. The Periodic Autocorrelation Function (PeACF) and the Periodic Partial Autocorrelation Function (PePACF) serve as useful indicators of the correlation or of the dependence between the values of the series so that they play an important role in model identification. The identification is based on the cut-off property of the Periodic Autocorrelation Function (PeACF). We derive an explicit expression for the asymptotic variance of the sample PeACF to be used in establishing its bands. Therefore, we will get in this study a new structure of the periodic autocorrelation function which depends directly to the variance that will derived to be used in establishing its bands for the PMA process over the cut-off region and we have studied the theoretical side and we will apply some simulated examples with R which agrees well with the theoretical results.

Keywords: Periodic Models, PARMA Model, Identification, Periodic Autocorrelation Function

1. INTRODUCTION

The time series has been found that many meteorological variables (such as rainfall, global temperature) are no stationary. The theory and practice of time series analysis have developed rapidly since the appearance in 1970 of the seminal work of (Box *et al.*, 2013).

We Know that the time series analysis and modeling is an important tool in many areas in our life like water resources. It is used for building mathematical models to generate synthetic hydrologic records to forecast, determine the likelihood, detect trends and shifts, and to interpolate missing data and extend records in hydrologic records. The statistical characteristics of hydrologic series are important deciding factors in the selection of the type of model. For example, in most cases known in nature, river

flows have significant periodic behavior in the mean, standard deviation and skewness. In addition to these periodicities, they show a time correlation structure which may be either constant or periodic, (Anderson and Vecchia, 1993; Bartlett, 1946).

Many macroeconomic time series display a trend and marked seasonal variation, while many variables in finance and marketing display seasonality but no trend. If there is a trend in the data, then often one is interested in examining the nature of this trend, as this can have implications for forecasting and for subsequent model building, (Iqelan, 2011).

The fundamental aim of periodic time series analysis is generally two-fold: To understand and identify the stochastic process that produced the observed series and in turn to forecast future values of a series from past values alone. The common procedure in modeling such periodic river flow series is first to standardize or filter

Corresponding Author: Hazem I. El Shekh Ahmed, Department of Mathematics, Quds Open University, Gaza Strip, Palestine

the series and then fit an appropriate stationary model to the reduced series, Hipel and McLeod (1994). Most estimation techniques depend on the assumption that the series is stationary. A special class of nonstationary time series has been defined by (Gladyshev, 1961), called periodically correlated time series (also known as cyclostationary time series). These time series are nonstationary, but have periodic means and covariances.

The major aim of this study is the PARMA model identification, i.e., the determination of the seasonally varying orders of the PARMA model, and to develop a practical computer program which performs model identification for any given actual periodic stationary series. An important class of periodic models useful in such situations consists of PARMA models, which are extensions of commonly used ARMA models that allow periodic parameters. PARMA models explicitly represent the seasonal fluctuations in mean flow, flow standard deviation, and flow autocorrelation, resulting in a more realistic time series model that leads to more reliable simulations of natural river flows, (Ula and Smadi 2003; Vecchia, 1985).

Since PARMA models are quite new, many questions about them are still unanswered and need further studies. It is known that there exists general methods for the identification of standard ARMA models of mixed type, however no such satisfactory method is available for PARMA processes. In the application part of this study, PARMA models may be more suitable for some seasons (**Table 1**). This study involves only the periodic stationarity case and the programs work only for periodic stationary processes. There exists well-known methods like differencing or filtering to achieve stationarity of standard ARMA models, such methods for achieving periodic stationarity should also be investigated.

Table 1. Behavior of the PeACF and PePACF for PARMA models, where the PeACF have values before lag q_s but it is zero for lags beyond q_s for pure PMA processes and the order of the process can be decided according to the sample PeACF. Also the PePACF have values before lag p_s but it becomes zero for lags beyond p_s for pure PAR processes and the order of the process can be decided according to the sample PePACF

PAR (p_s)	PMA (q_s)	PARMA (p_s, q_s)	
PeACF	Tails off	Cuts-off after lag q_s	Tails off
PePACF	Cuts-off after lag p_s	Tails off	Tails off

Of specific importance to this study is to establish an identification of a possible model based on an available realization, i.e., to decide the kind of the model with correct orders by using two statistics functions widely used for identifying PARMA time series models which are PeACF and PePACF. Also, we will get some properties of the variance summarized which are needed for the assessment of the cut-off property of the seasonal ACF for a season s which follows a $MA(q(s))$ and then apply this establish on some simulated examples.

2. STEPS FOR MODEL IDENTIFICATION

In time series analysis, the periodic autocorrelation function PeACF and the periodic partial autocorrelation function PePACF serve as useful indicators of the correlation or of the dependence between the values of the series so that they play an important role in model identification (Box *et al.*, 2013).

The most crucial steps are to identify and build a model based on available data (**Fig. 1**). This requires a good understanding of the processes, particularly the characteristics of these processes in terms of their PeACF and PePACF. In practice, these PeACF and PePACF are unknown, so they have to be estimated by the sample PeACF and PePACF.

Thus, in model identification our goal is to match patterns in the sample PeACF and PePACF for the PARMA models William (2006).

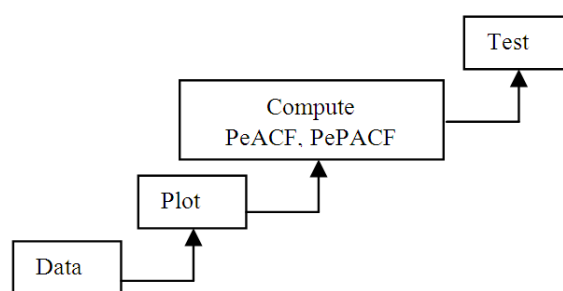


Fig. 1. Summary of steps for model identification which summarize in collecting data then plot the time series data then compute and examine the sample PeACF and the sample PePACF of the series and finally test the deterministic terms

- Step 1: Plot the time series data and choose proper transformation, where the most commonly used transformation are variance-stabilizing transformations
- Step 2: Compute and examine the sample PeACF and the sample PePACF of the original series to further confirm a necessary degree of them
- Step 3: Compute and examine the sample PeACF and the sample PePACF of the properly transformed series to identify the orders of $p = \max\{p_1, \dots, p_d\}$ and $q = \max\{q_1, \dots, q_d\}$ in PARMA_d ($p_1, q_1, p_2, q_2, \dots, p_d, q_d$)
- Step 4: Test the deterministic terms $\phi_{s,0}$ and $\theta_{s,0}$ when $d > 0$

Alternatively, one can include $\phi_{s,0}$ and $\theta_{s,0}$ initially and discard it at the final model estimation if the preliminary estimation result is not significant. This steps will applied in section of simulation with R program.

3. PEACF OF PARMA MODLS

We consider the identification of seasonally varying orders of PARMA time series models by making use of a periodic version of the functions ACF and PACF, to get a periodic autocorrelation function PeACF and periodic partial autocorrelation function PePACF.

Definition 3.1

Gladyshev (1961) We say that the process $\{X_t\}$ is a periodically correlated time series, if:

$$\exists d \in \mathbb{Z}^+; \exists \mu_t = \mu_{t+d}, \gamma_X(s, t) = \gamma_X(s + d, t + d) \forall s, t \in \mathbb{Z}$$

Where:

$$\mu_t = EX_t < \infty \text{ and } \gamma_X(s, t) = E[(X_s - EX_s)(X_t - EX_t)] < \infty$$

We will write the time index parameter t as $t(s, r) = s + rd$ such that $s = 1, \dots, d$ refers to season and $r + 1 (r = 0, \dots, N-1)$ refers to year, d is the number of seasons and N is the number of years with total sample size $n = Nd$.

Let $\{X_{s+rd}\}$ be a zero-mean periodically correlated time series of period $d = 2, 3, \dots$. The PARMA_d ($p_1, q_1, p_2, q_2, \dots, p_d, q_d$) has representation Equation 3.1:

$$X_{s+rd} - \sum_{i=1}^{p_s} \phi_{s,i} X_{s+rd-i} = \epsilon_{s+rd} - \sum_{i=1}^{q_s} \theta_{s,i} \epsilon_{s+rd-i} \quad (3.1)$$

where, $\{\epsilon_{s+rd}\}$ is an uncorrelated periodic white noise process and normally distributed terms with mean zero and

periodic variances $\sigma_{\epsilon}^2(s)$ and $\phi_{s,i}, \theta_{s,i}$ are the autoregression and moving average coefficients respectively.

For the univariate periodic stationary PARMA process $\{X_{s+rd}\}$, the Periodic Autocovariance Function (PeACVF) is defined as Equation 3.2:

$$\begin{aligned} \gamma_s(h) &= E(X_{s+rd} X_{s+rd-h}) \\ &= \begin{cases} \sum_{i=0}^{q_s-h} \theta_{s,i+h} \theta_{s-h,i} \sigma_{s-h-i}^2, & \text{if } 0 \leq h \leq q_s; \\ 0, & \text{if } h > q_s \end{cases} \end{aligned} \quad (3.2)$$

where, $\theta_{s,0}$ is defined to be 1 for all seasons s at lag $h \geq 0$. This is similar to the cut-off property of the ACVF of the MA processes.

The PeACF of $\rho_s(h)$ for season s at lag $h \geq 0$ is defined as Equation 3.3:

$$\begin{aligned} \rho_s(h) &= \begin{cases} \frac{\sum_{i=0}^{q_s-h} \theta_{s,i+h} \theta_{s-h,i} \sigma_{s-h-i}^2}{\sqrt{\gamma_s(0) \gamma_{s-h}(0)}}, & \text{if } 1 \leq h \leq q_s \\ 0, & \text{if } h > q_s. \end{cases} \\ &= \begin{cases} \frac{\sum_{i=0}^{q_s-h} \theta_{s,i+h} \theta_{s-h,i} \sigma_{s-h-i}^2}{\sqrt{\sum_{i=0}^{q_s} \theta_{s,i}^2 \sigma_{s-i}^2 \sum_{i=0}^{q_s} \theta_{s-h,i}^2 \sigma_{s-h-i}^2}}, & \text{if } 1 \leq h \leq q_s \\ 0, & \text{if } h > q_s \end{cases} \end{aligned} \quad (3.3)$$

where, $\gamma_s(0)$ is the variance for the s^{th} season.

This is an important property of the PeACF. That is, for a season following a pure MA(q_s) process, $\rho_s(h) = 0 \forall h > q_s$, which is called the cut-off property of the periodic autocorrelation function of a pure PMA processes. Also, this function does not play the same role with PARMA and PAR models.

Definition 3.2

The PeACF of $\{X_t\}$ for season s is defined as Equation 3.4:

$$\begin{aligned} \rho_s(h) &= E \left[\left(\frac{X_{s+rd} - \mu_s}{\sqrt{\gamma_s(0)}} \frac{X_{s+rd-h} - \mu_{s-h}}{\sqrt{\gamma_{s-h}(0)}} \right) \right] \\ &= E(Z_{s+rd} Z_{s+rd-h}) \\ &= \frac{\gamma_s(h)}{\sqrt{\gamma_s(0) \gamma_{s-h}(0)}} \quad h \geq 0 \end{aligned} \quad (3.4)$$

where, $\gamma_s(0)$ is the variance for the s^{th} season and $\{Z_{s+rd}\}$ denotes the periodically standardized time series.

We can define a function called *peacf* which summarizes the steps of periodic autocorrelation function PeACF to identify the seasonally varying orders of PARMA time series models. This function, which implements formula (3.3), is represented in Algorithm 1.

Algorithm 1. *peacf* of PeACF for PARMA_d(0; q_s) process

Input: $(\sigma_{s-h-i}^2, g_s(h)), s = 1, \dots, d$

for $s \leftarrow 1$ to d do

 for $i \leftarrow 0$ to $q_s - h$ do

$\rho_s(h) \leftarrow \begin{cases} \frac{\gamma_s(h)}{\sqrt{\gamma_s(0)\gamma_{s-h}(0)}}, & 1 \leq h \leq q_s; \\ 0, & h > q_s. \quad // \text{ cut-off} \end{cases}$

 end

end

Output: $\rho_s(h)$

Although, as for the ACF of a stationary ARMA model, it can be shown for a periodic stationary PARMA model that $\lim_{h \rightarrow \infty} \rho_h(s) = 0$ (i.e., PeACF of a stationary series goes to zero as time lag increases) (Ula and Smadi, 2003). The cut-off property of $\rho_h(s)$ for an arbitrary season s following the PMA $[q(s)]$ process Equation 3.5:

$$X_{s+rd} = \epsilon_{s+rd} - \sum_{i=1}^{q_s} \theta_{s,i} \epsilon_{s+rd-i} \quad (3.5)$$

where, is that $\rho_h(s) = 0$ for all $h > q(s)$. This is analogous to the cut-off property of the ACF of an ordinary MA process.

Definition 3.3

The sequences of $\{\psi_{s,i}\}$ is called absolutely summable sequences (i.e., converge) if $\sum_{i=0}^{\infty} |\psi_{s,i}| < \infty$.

Definition 3.4

The periodic autoregressive moving-average process for the time series $\{X_{s+rd}\}$ is said to be causal function of $\{\epsilon_{s+rd}\}$, if for each season $s = 1, 2, \dots, d$, there exist sequences of $f \{\psi_{s,i}\}$, such that Equation 3.6:

$$X_{s+rd} = \sum_{i=0}^{\infty} \psi_{s,i} \epsilon_{s+rd-i}, \text{ for each } s = 1, \dots, d; r = 0, 1, 2, \dots \quad (3.6)$$

where, $\psi_{s,0} = 1$ and all the values $\psi_{s,i}$ are d -periodic and satisfy $\sum_{i=0}^{\infty} |\psi_{s,i}| < \infty$ for each season s .

Proposition 3.1

Let $\{X_{s+rd}\}$ to be periodically correlated time series and the absolutely summable sequences $\{\psi_{s,i}\}$ be such

that $\{\psi_{s,i}\}$ is d -periodic. Then the autocovariance function $\gamma_s(h) = E[X_{s+rd}X_{s+rd-h}]$ is given by Equation 3.7:

$$\gamma_X(s, s-h) = \sum_{l=h}^{\infty} \sum_{r=0}^{d-1} \sum_{q=0}^{\infty} \psi_{s, h-l+r} \psi_{s-h, r} \gamma_{\epsilon}(s-h-r+l, s-h-r) \quad (3.7)$$

Proof

Let $\{X_{s+rd}\}$ be a zero-mean periodically correlated time series of period $d = 2, 3, \dots$.

Then:

$$\begin{aligned} E[X_{s+rd} X_{s+rd-h}] &= \lim_{n \rightarrow \infty} E \left[\sum_{i=0}^n \psi_{s,i} \epsilon_{s+rd-i} \right] \\ &= \lim_{n \rightarrow \infty} E \left[\sum_{i=0}^n \psi_{s-h, i} \epsilon_{s+rd-h-i} \right] \\ &= \lim_{n \rightarrow \infty} E \left[\sum_{i,j=0}^n \psi_{s,i} \psi_{s-h, j} \epsilon_{s+rd-i} \epsilon_{s+rd-h-j} \right] \\ &= \lim_{n \rightarrow \infty} \sum_{i,j=0}^n \psi_{s,i} \psi_{s-h, j} \gamma_{\epsilon}(s-i, s-h-j) \\ &= \sum_{i,j=0}^{\infty} \psi_{s,i} \psi_{s-h, j} \gamma_{\epsilon}(s-i, s-h-j) \end{aligned}$$

Since the autocovariance function of $\{\epsilon_{s+rd}\}$ is bounded by $\max. [\gamma_{\epsilon}(1, 1), \dots, \gamma_{\epsilon}(d, d)]$ and the elementwise product of absolutely summable sequences is absolutely summable.

Now, let us change the index i by $l = h+j-i$ to get:

$$\begin{aligned} E[X_{s+rd} X_{s+rd-h}] &= \sum_{l=h}^{\infty} \sum_{j=0}^{\infty} \psi_{s, h+j-l} \psi_{s-h, j} \gamma_{\epsilon}(s-h-j+l, s-h-j) \\ &= \sum_{l=h}^{\infty} \sum_{r=0}^{d-1} \sum_{q=0}^{\infty} \psi_{s, h-l+r+qd} \psi_{s-h, r+qd} \gamma_{\epsilon}(v+l-qd, v-qd) \end{aligned}$$

where, $j = r + qd$ and $v = s-h-r$.

Therefore:

$$\begin{aligned} E[X_{s+rd} X_{s+rd-h}] &= \sum_{l=h}^{\infty} \sum_{r=0}^{d-1} \left[\sum_{q=0}^{\infty} \psi_{s, h-l+r+qd} \psi_{s-h, r+qd} \right] \gamma_{\epsilon}(v+l, v) \end{aligned}$$

Corollary 3.1

If $\{\epsilon_{s+rd}\}$ in proposition (3.1) is periodic white noise PWN $(0, \sigma_s^2, d)$, then Equation 3.8:

$$\gamma_X(s, s-h) = \sum_{r=0}^{d-1} \psi_{s, h+r} \psi_{s-h, r} \sigma_{s-h-r}^2 \quad (3.8)$$

4. PEPACF OF PARMA MODELS

In time series analysis, the periodic partial autocorrelation function is well adapted to the identification of pure PAR processes. Let $\{X_{s+rd}\}$ be a zero-mean periodically correlated time series of period $d = 2, 3, \dots$. The $PAR_d(p_1, \dots, p_d)$ has representation Equation 4.1:

$$X_{s+rd} = \sum_{i=1}^{p_s} \phi_{s,i} X_{s+rd-i} + \epsilon_{s+rd} \quad (4.1)$$

Definition 4.1

The conditional expectation of X_t given $X_s, s < t$ is denoted by $E(X_t | X_s, s < t)$ and defined as follows $E(X_t | X_s, s < t) = a_1 X_{t-1} + a_2 X_{t-2} + \dots$.

The conditional expectation can be replaced by the linear projection. The best linear prediction of $X_t, s < t$ is denoted by $\text{pred}(X_t | X_s, s < t)$, Iqelan (2011).

Consider the discrete time series $X_{s+rd}, X_{s+rd+h}, X_{s+rd-(h-1)}, \dots, X_{s+rd-1}, X_{s+rd}, \dots$. Hence, for any $h > 1$, define $\hat{X}_{s+rd}^{(h-1)} = \text{Proj}(X_{s+rd} | X_{s+rd-1}, \dots, X_{s+rd-(h-1)})$ to be the best forward linear predictor of X_{s+rd} from the intermediate variables X_{s+rd-1} to $X_{s+rd-(h-1)}$, where the upper indices refer to quantities connected to the predictor of X_{s+rd} based on the $h-1$ observations right before or right after it. So Equation 4.2:

$$\hat{X}_{s+rd}^{(h-1)} = \sum_{i=1}^{h-1} a_{h-1,i} X_{s+rd-(h-i)} \quad (4.2)$$

where, $a_{h-1,i}, i = 1, \dots, h-1$ are the forward coefficients of the prediction.

Let $\epsilon_{s+rd}^{(h-1)} = X_{s+rd} - \hat{X}_{s+rd}^{(h-1)}$ to be the forward residual of X_{s+rd} and define Equation 4.3:

$$\text{Var}(\epsilon_{s+rd}^{(h-1)}) = {}^f \sigma_{s+rd}^2(h-1) \quad (4.3)$$

We can also obtain the best backward linear prediction if we reversing the time index, say $\tilde{X}_{s+rd-h}^{(h-1)} = \text{Proj}(X_{s+rd-h} | X_{s+rd-(h-1)}, \dots, X_{s+rd-1})$ to be the backward linear predictor of X_{s+rd-h} from $X_{s+rd-(h-1)}$ to X_{s+rd-1} . So Equation 4.4:

$$\tilde{X}_{s+rd-h}^{(h-1)} = \sum_{i=1}^{h-1} b_{h-1,i} X_{s+rd-i} \quad (4.4)$$

where, $b_{h-1,i}, i = 1, \dots, h-1$ are the backward coefficients of the prediction.

Also, let $\eta_{s+rd-h}^{(h-1)} = X_{s+rd-h} - \tilde{X}_{s+rd-h}^{(h-1)}$ to be the backward residual of X_{s+rd-h} and define Equation 4.5:

$$\text{Var}(\eta_{s+rd-h}^{(h-1)}) = {}^b \sigma_{s+rd}^2(h-1) \quad (4.5)$$

Definition 4.2

Iqelan (2011) The periodic partial autocorrelation function $\beta(s, s-h)$ is the correlation between X_{s+rd} and X_{s+rd-h} with the effect of the intermediate variables $X_{s+rd-1}, \dots, X_{s+rd-(h-1)}$ "filtered out", which is defined on $\mathbb{Z} \times \mathbb{Z}$ by Equation 4.6:

$$\beta(s, s-h) = \begin{cases} \text{Var}(X_{s+rd}), & \text{if } h=0; \\ \text{Corr}(X_{s+rd}, X_{s+rd-h}), & \text{if } h=1; \\ \text{Corr}(X_{s+rd} - \hat{X}_{s+rd}^{(h-1)}, X_{s+rd-h} - \tilde{X}_{s+rd-h}^{(h-1)}), & \text{if } h \geq 2. \end{cases} \quad (4.6)$$

Notice that, setting $\beta(s, s) = \text{Var}(X_{s+rd})$ instead of 1 in the above definition. Also, we can rewrite $\beta(s, s-h)$ at lag h of one variable to be as Equation 4.7:

$$\begin{aligned} \beta_s(h) &= \beta(s, s-h) \\ &= \text{Corr}(X_{s+rd} - \hat{X}_{s+rd}^{(h-1)}, X_{s+rd-h} - \tilde{X}_{s+rd-h}^{(h-1)}) \\ &= \frac{\text{Cov}(\epsilon_{s+rd}^{(h-1)}, \eta_{s+rd-h}^{(h-1)})}{\sqrt{{}^f \sigma_{s+rd}^2(h-1)} \sqrt{{}^b \sigma_{s+rd}^2(h-1)}} \end{aligned} \quad (4.7)$$

Theorem 4.1:

- If $\{X_{s+rd}\}$ is a causal periodic autoregression of order p_s at season s , then $\beta_s(h) = 0$ whenever $h > p_s$
- If $\{X_{s+rd}\}$ is a periodic series with period d and $\beta_s(h) = 0$ for all $h > p_s$ and $\beta_s(p_s) \neq 0$ for each season s , then $\{X_t\}$ is a periodic autoregression of order p_s at season s ; $1 \leq s \leq d$

It was mentioned that for a pure PMA process, the PeACF is zero for lags beyond q_s . Likewise, for pure PAR processes, the PePACF becomes zero for lags beyond p_s , so if the correct order is p_s for season s , then $\beta_s(h) = 0$ for all $h > p_s$. This is the cut-off property of periodic partial autocorrelation function for PAR processes.

Actually, the PePACF for PMA models behaves much like the PeACF for PAR models.

Also, the PePACF for PAR models behaves much like the PeACF for PMA models.

Also, we can define a function called pepacf which summarizes the steps of periodic partial autocorrelation function PePACF to identify the seasonally varying orders of PARMA time series models which implements formula (4.7) to be represented in Algorithm 2.

Algorithm 2. pepacf of PePACF for $PARMA_d(p_s; 0)$ process

Input:

$$(\sigma_{s+rd}^2(h-1), \sigma_{s+rd}^2(h-1), \epsilon_{s+rd}^{(h-1)}, \eta_{s+rd-h}^{(h-1)}, \text{Cov}(\epsilon_{s+rd}^{(h-1)}, \eta_{s+rd-h}^{(h-1)})).$$

for $s \leftarrow 1$ to d do

$$\beta_s(h) \leftarrow \begin{cases} \text{Corr}(X_{s+rd} - \hat{X}_{s+rd}^{(h-1)}, X_{s+rd-h} - \tilde{X}_{s+rd-h}^{(h-1)}), & 1 \leq h \leq p_s \\ 0, & h > p_s \end{cases}$$

// cut-off

end

Output: $\beta_s(h)$.

5. SAMPLE PERIODIC AUTOCORRELATION FUNCTION

Assume that $\{X_1, X_2, \dots, X_{Nd}\}$ be a series of size Nd from a periodic stationary process $\{X_{s+rd}\}$. Then $\hat{\gamma}_h(s)$ is the sample periodic autocovariance function which calculated from Equation 5.1:

$$\begin{aligned} \hat{\gamma}_h(s) &= \frac{1}{N} \sum_{r=0}^{N-1} \left(X_{s+rd} - \frac{1}{N} \sum_{r=0}^{N-1} X_{s+rd} \right) \\ &\quad \left(X_{s+rd-h} - \frac{1}{N} \sum_{r=0}^{N-1} X_{s+rd-h} \right) \quad (5.1) \\ &= \frac{1}{N} \sum_{r=0}^{N-1} (X_{s+rd} - \bar{X}_s) (X_{s+rd-h} - \bar{X}_{s-h}) \end{aligned}$$

where, the sample mean for season s is $\bar{X}_s = \frac{1}{N} \sum_{r=0}^{N-1} X_{s+rd}$.

Also, the sample periodic autocorrelation function can be given by Equation 5.2:

$$r_h(s) = \frac{\hat{\gamma}_h(s)}{\sqrt{\hat{\gamma}_0(s)\hat{\gamma}_0(s-h)}}, h \geq 0 \quad (5.2)$$

Since μ_s and $\gamma_h(s)$ are periodic with period d , also \bar{X}_s and $\hat{\gamma}_h(s)$ are so.

Let the white noise terms to be independent and normal, so that $\{X_{s+rd}\}$ is a Gaussian PARMA process,

then in (Pagano, 1978) it is proved that $\hat{\gamma}_h(s)$ are consistent, asymptotically independent, joint normal, unbiased estimates, efficient and converge almost surely to $\gamma_h(s)$ for all s and h . Also, \bar{X}_s is consistent and unbiased estimator of μ_s and it can be shown that it also consistent under the periodic stationarity assumption.

For stationary processes, the asymptotic joint normality and unbiasedness of the sample ACF $r_h(s)$ with $d = s = 1$ have been shown by (Bartlett, 1946) and the asymptotic variancecovariance matrix has been specified in which Equation 5.3:

$$\begin{aligned} \text{Var}[r_h(s)] &\cong \\ \frac{1}{N} \sum_{m=-\infty}^{\infty} \{ &\rho_m^2 + \rho_{m+h}\rho_{m-h} - 4\rho_h\rho_m\rho_{m-h} + 2\rho_m^2\rho_h^2 \} \end{aligned} \quad (5.3)$$

where, the symbol “ \cong ” wherever appears means that the statement is true for large N (Box *et al.*, 2013).

In Vecchia (1985) an approximate solution for the first- and second-order moments of $r_h(s)$ were obtained by pretending that the sample means \bar{X}_s and variances $\hat{\gamma}_0(s)$ in $r_h(s)$ are equal to their population counterparts μ_s and $\gamma_h(s)$, respectively. This assumption will obviously be well justified only for large samples due to the consistency of these estimators. In that case, we can take $r_h(s)$ as Equation 5.4:

$$r_h(s) = \frac{1}{N} \sum_{r=0}^{N-1} Z_{s+rd} Z_{s+rd-h} \quad (5.4)$$

It then follows that $r_h(s)$ is asymptotically unbiased with asymptotic variance Equation 5.5:

$$\begin{aligned} \text{Var}[r_h(s)] &\cong \\ \frac{1}{N} \sum_{m=-\infty}^{\infty} [&\rho_{md}(s)\rho_{md}(s-h) + \rho_{md+h}(s)\rho_{md-h}(s-h)] \end{aligned} \quad (5.5)$$

Which by utilizing $\rho_s(h) = \rho_{s-h}(-h)$ and the fact that $\rho_{md}(s)\rho_{md}(s-h) + \rho_{md+h}(s)\rho_{md-h}(s-h)$ is an even function of m , reduces to Equation 5.6:

$$\begin{aligned} \text{Var}[r_h(s)] &\cong \frac{1}{N} \\ \left(1 + [\rho_h(s)]^2 + 2 \right. & \left. \left\{ \sum_{m=1}^{\infty} \rho_{md}(s)\rho_{md}(s-h) + \rho_{md+h}(s)\rho_{md-h}(s-h) \right\} \right) \end{aligned} \quad (5.6)$$

Now, the Equation 5.5 and 5.6 is used, only the first two terms are retained in these Equation. The remaining last two terms containing the third-and higher-order autocorrelations disappear for $h>q(1)$ if the first season follows a $MA(q(1))$ process. Therefore, for the assessment of cut-off in seasonal autocorrelation, that is for checking whether season s is a $MA(q(1))$ or not, for example, for which $\rho_h(1) = 0$, for $h>q(1)$.

In the following proposition some properties of the variance of $Var[r_h(s)]$, which follow from (5.6), are summarized which are needed for the assessment of the cut-off property of the seasonal ACF for a season s which follows a $MA(q(s))$ process.

Proposition 5.1

Ula and Smadi (2003) Let $\{X_{s+rd}\}$ be a periodic stationary $PARMA_d(p(s), q(s))$ process.

If s is an arbitrary with $p(s) = 0$, then, for positive integer h , we have the following results for $Var[r_h(s)]$;

(i) for $q(s)<d$ Equation 5.7:

$$Var[r_h(s)] \cong \begin{cases} \frac{1}{N}(1+[\rho_h(s)]^2), & h \leq q(s) \\ \frac{1}{N}, & h > q(s) \end{cases} \quad (5.7)$$

(ii) for $rd \leq q(s) < (r+1)d$, $r = 1, 2, \dots$, then:

$$Var[r_h(s)] \cong \begin{cases} \frac{1}{N}(1+[\rho_h(s)]^2 + 2\{\sum_{m=1}^r \rho_{md}(s)\rho_{md}(s-h) + \rho_{md+h}(s)\rho_{md-h}(s-h)\}), & 1 \leq h \leq q(s) - rd \\ \frac{1}{N}(1+[\rho_h(s)]^2 + 2\{\sum_{m=1}^r \rho_{md}(s)\rho_{md}(s-h) + \sum_{m=1}^j \rho_{md+h}(s)\rho_{md-h}(s-h)\}), & q(s) - (j+1)d < h \leq q(s) - jd, j = 1, \dots, (r-1) \\ \frac{1}{N}(1+[\rho_h(s)]^2 + \{\sum_{m=1}^r \rho_{md}(s)\rho_{md}(s-h)\}), & q(s) - d < h \leq q(s) \\ \frac{1}{N}(1+2\sum_{m=1}^r \rho_{md}(s)\rho_{md}(s-h)), & h > q(s) \end{cases}$$

Proof

For more details Ula and Smadi (2003).

Setting $d = s = 1$, it can be easily seen from Proposition (5.1) that, for $h>q(s) = q$, case (i) reduces to the white noise process Equation 5.8:

$$Var[r_h(s)] = \frac{1}{N}, h \geq 1 \quad (5.8)$$

Also, case (ii) reduces to the $MA(h)$ process Equation 5.9:

$$Var[r_h(s)] \cong \frac{1}{N} \left(1 + 2 \sum_{m=1}^r \rho_m^2 \right), h > r \quad (5.9)$$

These are the well-known formulas for the identification of white noise and MA processes, respectively, in the context of stationary processes (Box *et al.*, 2013). However, note that for $h \leq q(s)$ the formulas for $Var(r_h)$ in case (ii) of Proposition (4.1) are rather approximate as they are base on (5.6).

The identification bands for the $MA(q(s))$ process utilize $r_h(s)$ being asymptotically normal with zero mean and variance Equation 5.8 and 5.9 over the cut-off region $h>q(s)$.

Following the same methodology applied to the sample ACF of a stationary process, for a season s which follows a $MA(q(s))$ process, we start checking values of $q(s)$ successively, starting with $q(s) = 1$. Then for large N , as long as $q(s)<d$, Equation 5.8 implies that $r_h(s)$, for $h>q(s)$, is normally distributed with mean zero and

variance $1/N$, so that the 95% band $\left(\frac{-1.96}{\sqrt{N}}, \frac{1.96}{\sqrt{N}} \right)$ is

applied to those autocorrelations. If $q(s) \geq d$, which is unlikely for moderate or large values of d , Equation 5.9 should be utilized, with $\rho_h(s)$ estimated by $r_h(s)$. let s_h to denoted the sample value of $\sqrt{Var[r_h(s)]}$, the 95% band is $(-1.96 s_h, 1.96 s_h)$, which should be applied for $r_h(s)$, for $h>q(s)$, the accuracy of these bands are verified through simulation next section.

6. SIMULATION RESULT

In this section we will illustrate some simulation studies to investigate the usefulness in practice of the theoretical results stated earlier and to demonstrate the identification procedure. The *R* programming language was used in this simulation study in conjunction with the *pear* package.

Example 6.1

Consider the model $PARMA_4(0, 1; 2, 2; 3, 0; 0, 4)$:

$$\begin{aligned}
 X_{1+4r} &= 0.9\epsilon_{4+4(r-1)} + \epsilon_{1+4r} \\
 X_{2+4r} &= 0.7X_{1+4r} + 0.8X_{4+4(r-1)} \\
 &\quad + 0.6\epsilon_{1+4r} + 1.2\epsilon_{4+4(r-1)} + \epsilon_{2+4r} \\
 X_{3+4r} &= 1.1X_{2+4r} - 0.6X_{1+4r} \\
 &\quad + 0.5X_{4+4(r-1)} + \epsilon_{3+4r} \\
 X_{4+4r} &= 0.5\epsilon_{3+4r} + 0.7\epsilon_{2+4r} \\
 &\quad + 0.8\epsilon_{1+4r} + 0.3\epsilon_{4+4(r-1)} + \epsilon_{4+4r}
 \end{aligned}$$

Which is chosen to be periodic stationary and invertible for determining periodic stationarity and invertibility of such processes, the first season MA (1), second season ARMA (2,2), third season AR (3) and fourth season MA (4).

To process our simulation study an R-code was written to generate 10000 independent replicates. Each of these replicates has sizes $N = 2500$, $N = 5000$ and $N = 7500$.

In all cases, the white noise terms are independently and normally distributed with mean zero and variance equal to one. In **Fig. 2** we see that one of simulated series which corresponds to $N = 10000$.

The white noise terms are assumed to be independently and normally distributed with zero means (our basic assumptions) and unit variances ($\sigma_a^2(s) = 1; s = 1, 2, 3, 4$). Ten thousand realizations each of length N (years), i.e., $4 \times N$ values, for $N = 2500$, 5000 , 7500 , are simulated from the above 4-period PARMA model. For each realization, the sample PeACF, $r_h(s)$, for $s = 1, 2, 3, 4$ and $h = 1, \dots, 10$ are computed in **Table 2**. Only the first and fourth seasons here follow a pure MA process, MA (1) and MA (4), respectively. For the first season, we expect a

cut-off behavior for the sample PeACF for $h > 1$ and for the fourth season for $h > 4$. These can be observed from the values in **Table 2**. Only for these first seasons, we observe an apparent sudden drop in the values, for more analysis see **Fig. 3**.

Equation 5.8 follows and the relative frequencies (rel. freq.) of $r_h(1)$, $h > q = 1$, going outside the 95% band $\left(\frac{-1.96}{\sqrt{N}}, \frac{1.96}{\sqrt{N}}\right)$ for $N = 2500; 5000; 7500$, are given in

Table 2 respectively and shown in **Fig. 4**, which agree well with the theoretical asymptotic value of 5%. The relative frequency is the percentage of autocorrelation values over all realizations falling outside the corresponding bands. The asymptotic property also justifies the improvement in the values as N increases.

For the fourth season, therefore Equation 5.9 applies with relative frequencies of $r_h(4)$ going outside the corresponding 95% band for the three N values.

Example 6.2

In this example we will to see the effect of variance on the results, we take the same model in the previous example $PARMA_4(0, 1; 2, 2; 3, 0; 0, 4)$ and the white noise terms are independently and normally distributed with mean zero and variance equal 0.5 (possible take any value of the standard deviation).

We will see the variances seem to have no significant effect on the overall result, as theoretically expected. The simulation is repeated for $N = 10000$ and $\sigma_a^2(s) = 0.5$ for $s = (1, 2, 3, 4)$. For the first season, we expect a cut-off behavior for the sample PeACF for $h > 1$ and for the fourth season for $h > 4$. These can be observed from the values in **Table 3**

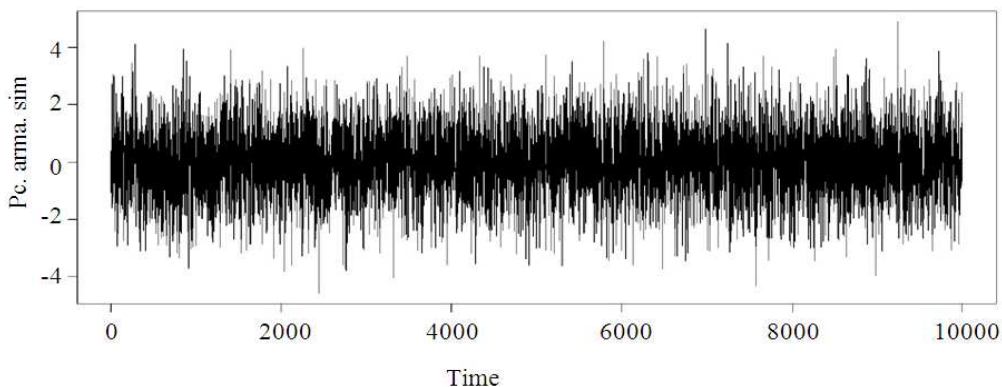


Fig. 2. Line graph of 10000 observations

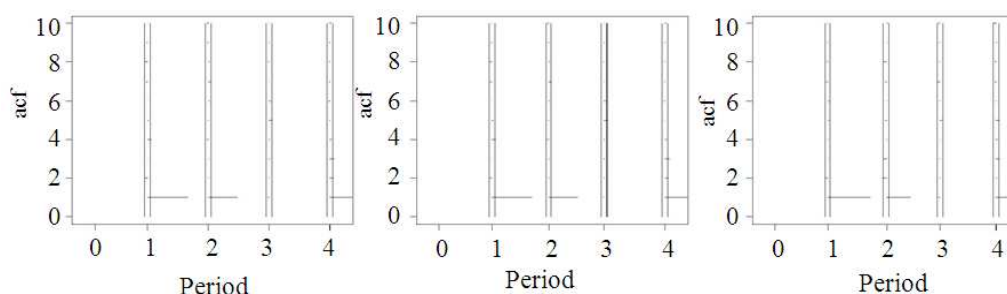


Fig. 3. The PeACF for the size data sets ($N = 2500$, $N = 5000$ and $N = 7500$ respectively). In season one PeACF show cut-off after lag 1 and since PeACF shows better cut-off, this season is said to follow MA (1) model. Also, in season four PeACF show cut-off after lag 4 and this season is said to follow MA (4) model.

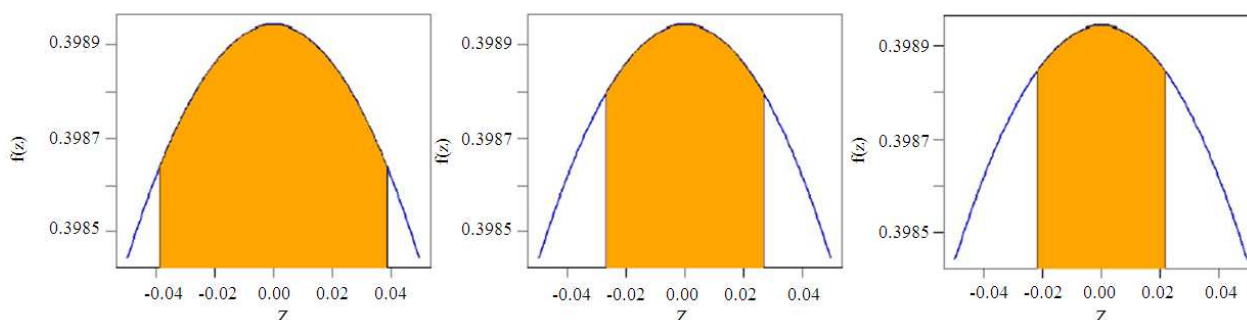


Fig. 4. The 95% confidence interval size data sets ($N = 2500$, $N = 5000$ and $N = 7500$ respectively). It can be said that the first, second, third and the fourth seasons (for $N = 2500$) follow white noise process without any doubt, since all values fall inside the bands for PeACF, but for season one, 1st lag in PeACF is outside the bands and since that value is close to band limits, it may be ignorable. Also, the seasons (for $N = 5000$) follow white noise process, but for season one and four, 1st lag in PeACF may be ignorable. Finally, the seasons (for $N = 7500$) follow white noise process, but for season one, two and four, 1st lag in PeACF may be ignorable.

Table 2. Average sample PeACF $r_h(s)$ for lag $h = 10$ with $\sigma_a^2 = 1$

h\s	$N = 2500$				$N = 5000$				$N = 7500$			
	1	2	3	4	1	2	3	4	1	2	3	4
1	0.605	0.378	0.000	0.430	0.592	0.388	0.232	0.4370	0.598	0.384	-0.003	0.455
2	-0.003	0.008	-0.031	0.020	-0.002	-0.007	0.647	-0.0320	0.000	-0.044	-0.606	-0.009
3	0.006	-0.005	0.000	-0.020	0.002	-0.018	-0.104	-0.0209	-0.001	-0.058	-0.126	0.000
4	-0.03	-0.002	-0.002	0.039	-0.007	-0.067	-0.199	-0.0270	-0.002	-0.040	0.000	0.003
5	0.019	-0.005	-0.023	0.036	-0.015	0.009	0.052	-0.0070	0.008	-0.006	0.007	-0.000
6	0.003	-0.019	0.008	-0.005	0.033	-0.001	-0.261	0.0080	-0.001	0.220	0.132	-0.001
7	0.016	0.013	0.004	0.001	-0.001	0.023	-0.048	-0.0150	0.002	-0.003	0.007	-0.019
8	-0.014	0.011	0.018	0.004	0.041	-0.009	0.012	-0.0110	0.008	0.030	0.022	0.001
9	-0.021	0.030	-0.008	0.003	-0.001	0.391	-0.007	0.0040	0.001	-0.001	0.000	0.009
10	-0.022	0.008	-0.008	0.000	0.005	0.040	0.021	-0.0060	0.000	0.041	0.007	0.008
Rel. freq.	1.220	1.580	1.320	0.680	0.240	0.420	0.520	0.8500	0.680	0.590	0.790	0.290

Table 3. Average sample PeACF $r_h(s)$ for lag $h = 10$ with $\sigma_a^2 = 0.5$. The variance seem to have no significant effect on the overall result, as theoretically expected

$N = 10000$				
h\s	1	2	3	4
1	0.612	0.370	-0.002	0.436
2	-0.005	0.003	-0.013	0.010
3	0.002	-0.008	0.014	-0.015
4	-0.001	-0.007	0.006	0.010
5	-0.000	0.014	-0.010	0.001
6	0.015	0.003	-0.006	-0.002
7	-0.001	0.016	-0.011	0.004
8	0.006	-0.000	0.002	0.002
9	0.005	0.007	0.007	0.003
10	0.001	0.015	0.003	-0.000

7. CONCLUSION

In this study we have made a survey on one of the most important topics in identification of PARMA models. An explicit expression for the asymptotic variance of the sample process PeACF is derived to be used in establishing its bands for the PMA process over the cut-off region and we have studied the theoretical side therefore we have some applications on it where the simulation results agree well with the theoretical results. In the future research, we will explicit expression for the asymptotic variance of the sample process PePACF that derived to be used in establishing its bands for the PAR process over the cut-off region and therefore we have some applications on it. Also, we will get estimation of PARMA models using some ways of estimation.

8. ACKNOWLEDGMENT

The researchers would like to thank the referee for carefully reading the paper and for helpful suggestions which greatly improved the paper.

9. REFERENCES

Anderson, P.L. and A.V. Vecchia, 1993. Asymptotic results for periodic autoregressive moving-average processes. *J. Time Series Anal.*, 14: 1-18. DOI: 10.1111/j.1467-9892.1993.tb00126.x

Bartlett, M.S., 1946. On the theoretical specification and sampling properties of autocorrelated time-series. *Suppl. J. Royal Stat. Society*, 8: 27-41.

Box, G.E.P., G.M. Jenkins and G.C. Reinsel, 2013. *Time Series Analysis: Forecasting and Control*. 4th Edn., John Wiley and Sons, Hoboken, ISBN-10: 1118619064, pp: 746.

Gladyshev, E.G., 1961. Periodically and almost-periodically correlated random processes with a continuous time parameter. *Theory Probab. Applied Math.*, 8: 173-177. DOI: 10.1137/1108016

Hipel, K.W. and A.I. McLeod, 1994. *Time Series Modelling of Water Resources and Environmental Systems*. Developments in Water Science. 1st Edn., Elsevier, Amsterdam, ISBN-10: 0080870368, pp: 1012.

Iqelan, B., 2011. *Periodically Correlated Time Series: Models and Examples*. 1st Edn., Lambert Academic Publishing, Saarbrücken, ISBN-10: 3844301100, pp: 204.

Pagano, M., 1978. On periodic and multiple autoregressions. *Annals Statist.*, 6: 1310-1317. DOI: 10.1214/aos/1176344376

Ula, T.A. and A.A. Smadi, 2003. Identification of periodic moving-average models. *Commun. Stat. Theory Meth.*, 32: 2465-2475. DOI: 10.1081/STA-120025388

Vecchia, A.V., 1985. Periodic Autoregressive-Moving Average (PARMA) modeling with applications to water resources. *J. Am. Water Resources Assoc.*, 21: 721-730. DOI: 10.1111/j.1752-1688.1985.tb00167.x

William, W.S.W., 2006. *Time Series Analysis: Univariate and Multivariate Methods*. 2st Edn., Pearson Addison Wesley, Boston, MA, ISBN-10: 0321322169, pp: 614.