Quantitative Estimation of Periodicities in Lyari River Flow Routing

Rana Khalid Naeem, Asif Mansoor

***Abstract*—**The hydrologic time series data display periodic structure and periodic autoregressive process receives considerable attention in modeling of such series. In this communication long term record of monthly waste flow of Lyari river is utilized to quantify by using PAR modeling technique. The parameters of model are estimated by using Frances & Paap methodology. This study shows that periodic autoregressive model of order 2 is the most parsimonious model for assessing periodicity in waste flow of the river. A careful statistical analysis of residuals of PAR (2) model is used for establishing goodness of fit. The forecast by using proposed model confirms significance and effectiveness of the model.

***Keywords*—**Diagnostic checks, Lyari river, Model selection, Monthly waste flow, Periodicity, Periodic autoregressive model.

I. Introduction

The exact mathematical models of hydrologic time series are never known. The inferred population models are only approximation. Estimation of models and their parameters from available data are often referred to in literature as time series modeling or stochastic modeling of hydrologic series. Now a day's development of time series modeling in hydrology has reached some degree of sophistication. It is an important tool for building models to determine the likelihood of extreme events, to forecast hydrologic events, to detect trends and shifts in stream flow records, and to fill in missing data and extend records. [1-2]. In hydrologic process it has been observe that these processes usually have seasonal mean, variance, skewness and serial dependence structure. Such seasonal streamflow series can be deseasonalized by subtracting the seasonal mean and then by dividing the seasonal standard deviation. By such procedures seasonality is only removes from mean and variance but it remains in serial dependence structure. To model such events, there are various approaches in the literature. The first approach is that of Box and Jenkins [3] which relies on moving average models for double differenced time series; so called seasonal ARIMA (SARIMA). Another approach is that of Harvey [4] which assumes that a seasonal time series can be decomposed into trend, cycle, seasonal and irregular components. Reduced forms of the resultant models are not much different from SARIMA models. The third approach is based on Hylleberg et al [5] that questions the adequacy of the double differencing filter in SARIMA models and mainly addresses the issue of how many unit roots should be imposed in autoregressive models. Finally, a fourth approach assumes the seasonal variation is best described by allowing the parameters in an autoregression to vary with the seasons that are called periodic autoregression (PAR) modeling. A periodic autoregression extends a non-periodic autoregressive model by allowing the autoregressive parameters to vary with seasons. A PAR model assumes that the observations in each of the season might be described by a different model. Periodic models are ideal for modeling hydrological time series since they are often periodically stationary and are now very much used in other disciplines like environmetric, Macroeconomics etc [6]. In this communication historical waste flow data from period January 1974 to December 2006 is fitted using periodic autoregressive approach. The parameters of most parsimonious model have been estimated and diagnostic checks have also been applied to examine the goodness of fit. The study established that the waste flow of the Lyari river follows PAR model of order 2. Forecasting using these estimates will be of use to managerial authorities for taking remedial measure to control the pollution caused by the river to the coastal water which have adverse effect on marine organisms such as phytoplankton, zooplankton and fish.

Rana Khalid Naeem is Professor of Mathematics University of Karachi, Karachi -75270 (e-mail: 2rknaeem@gmail.com).
Asif Mansoor is research scholar Department of Mathematics University of Karachi, Karachi-75270, Pakistan (Corresponding author: phone: 92-301-2990882; e-mail: asifmansure@yahoo.com).

II. Mathematical Formulation of PAR (p) Model

Consider a stochastic process, x_t having periodic structure observed monthly for N years so that $t = 1,2,3,\ldots,n$ $(=N / 12)$. A periodic autoregressive model of order p for the process can be represented as

$$x_t = I_s + \phi_{1s} x_{t-1} + \ldots + \phi_{ps} x_{t-p} + \varepsilon_t \qquad (1)$$

with $\varepsilon_t \sim iid(0, \sigma_\varepsilon^2)$, where $s = 1, 2, 3, \ldots\ldots, 12$ represents seasons and I_s is a seasonally–varying intercept terms. The ϕ_{is} are periodic autoregressive parameter for seasons s at lag i in the model, where $i = 1,2,3,\ldots, p$ which varies with the seasons for each lag.

In general, the PAR (p) process described by (1) can be written as an AR (p) model, for the twelve dimensional vector process (i.e. for each season). The model given by (1) can be rewritten as vector of monthly representation as

$$\Phi_0 X_T = I + \Phi_1 X_{T-1} + ...+ \Phi_P X_{T-P} + \varepsilon_T \tag{2}$$

where $X_T = (X_{1T}, X_{2T}, X_{3T}, ..., X_{12T})'$, $T=1,2,3...,N$, where X_{sT} is observation in season s in year T, $I = (I_1, I_2, I_3, I_4, ..., I_{12})'$, $\varepsilon_T = (\varepsilon_{1T}, \varepsilon_{2T}, \varepsilon_{3T}, ..., \varepsilon_{12T})'$ and P = $1+[(p-1)/12]$. The $\Phi_0, \Phi_1, ..., \Phi_P$ are (12x12) parameters matrices with elements

$$\Phi_0(i,j) = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } j > i \\ -\phi_{i-j,i} & \text{if } j < i \end{cases} \tag{3}$$

$$\Phi_k(i,j) = \phi_{i+12k-j,i}(i,j) \tag{4}$$

for i, j = $1,2,3,4,......, 12$ and $k = 1,2,3...,P$. The lower triangular of Φ_0 shows that (2) is in fact a recursive set of equation. For example, periodic autoregressive model of order 2 can be written as:

$$x_t = \alpha_{1s} x_{t-1} + \alpha_{2s} x_{t-2} + \varepsilon_t \tag{5}$$

whose multivariate representation is

$$\Phi_0 X_T = \Phi_1 X_{T-1} + \varepsilon_T \tag{6}$$

with

$$\Phi_0 = \begin{bmatrix} 1 & 0 & \cdots & & 0 \\ -\phi_{[1,2]} & 1 & 0 & & \vdots \\ -\phi_{[2,3]} & -\phi_{[1,3]} & 1 & \ddots & \\ \vdots & \vdots & & & 0 \\ -\phi_{[11,12]} & -\phi_{[10,12]} & \cdots & -\phi_{[1,12]} & 1 \end{bmatrix} \tag{7}$$

and

$$\Phi_1 = \begin{bmatrix} \phi_{[12,1]} & \cdots & \phi_{[2,1]} & \phi_{[1,1]} \\ 0 & \phi_{[12,2]} & & \phi_{[1,2]} \\ \vdots & 0 & \ddots & \vdots \\ 0 & 0 \cdots & 0 & \phi_{[12,12]} \end{bmatrix} \tag{8}$$

III. SELECTION OF ORDER

There exist many approaches in literature for model identification i.e. Box and Jenkins employed autocorrelation and partial autocorrelation in selecting order of model. Hurd and Gerr [7] and Bloomfield et al. [8] used graphical approach in deciding the order. Bentarzi and Hallins [9] proposed methods based on rank statistics for deciding order of model based on time series. All these approaches use different criteria for the selection of order of model. In the present study, Akiake's [10] and Schwarz [11] criterias are used for the selection of PAR (p) model. Akiake's information criteria is computed by utilizing (9)

$$\text{AIC } (p) = n \log \hat{\sigma}^2 + 24p \tag{9}$$

whereas Schwarz's criteria is by (10)

$$\text{BIC } (p) = n \log \hat{\sigma}^2 + 12 p \log n \tag{10}$$

where $\hat{\sigma}^2$ is the residual sum of square divided by the effective sample size n.

IV. DIAGNOSTICS FOR THE FITTED PAR (P) MODEL

Since residuals are unobservable that why is used to check model assumptions or adequacy of the model. The residuals ε_t for PAR (p) model are computed through equation

$$\varepsilon_t = x_t - \hat{x}_t \tag{11}$$

V. APPLICATION OF PAR (P) MODEL

Karachi, the provisional capital of sindh is located at the extreme west end of the Indus delta between north latitude 24° 51′ and east longitude 67° 4′. Lyari is one of the three rivers, along with the Malir and Hub rivers which flow through the greater metropolitan area of Karachi. The rivers watershed area has approximately 700 km^2 drainage area. It has approximately 200 km^2 within Karachi. Orangi and Gujar Nullahs are its main tributaries. These tributaries contribute two-third of the runoff within Karachi from the northwest to river. It carries the water that is purely combination of domestic sewage and industrial effluents. These effluents have very high load of pollutants which discharges into Arabian Sea.

The data are obtained from daily discharge measurements at the mouth of Arabian Sea, in cubic meter per second averaged over each of the respective months to obtain the monthly series. A partial time plot of the series is given in Fig. 1 which shows the nonstationarity in waste flow of river.

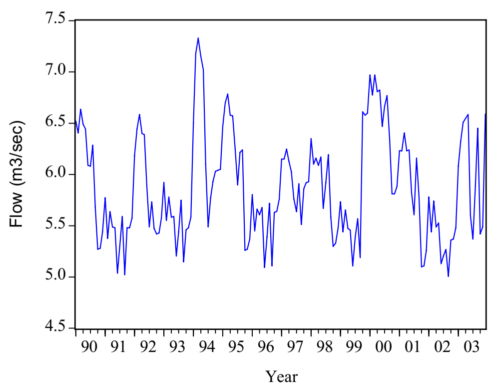

Fig. 1 Partial plot of monthly Lyari river flow

Now, first of all identify the order of the periodic autoregressive model which needs to obtain for estimation of parameters. For this purpose, Akaike Information Criterion (AIC) and Schwarz Criterion (SC) are computed by using equations (9) and (10) and F-statistics with their probabilities is calculated for the sets of the parameters of order *p+1* equal to zero. By using AIC & BIC information criteria and F-statistics, order of the suitable PAR model is selected. The Table I below provides the details of computed statistics, Akaike Information Criterion (AIC) and Schwarz Criterion (SC) with their probabilities values for the periodic autoregressive model of the order ranging between one and twelve.

The output of the test reported in Table I suggest that periodic autoregressive parameter of order two are significant for the waste flow of the Lyari river. Therefore, the parameter of PAR (2) model will be worked out.

TABLE 1
PERIODIC VARIATION TEST OUTPUT

PAR Order	AIC	BIC	F($\phi_{p+1,s} = 0$)	Prob. Value
1	-1138.6	-1139.5	3.5514814	.00000
2	-1192.7	-1193.2	0.2867891	.46912
3	-1109.3	-1110.6	1.3479240	.18987
4	-1052.0	-1053.8	0.9785696	.99107
5	-990.1	-992.3	4.7103478	.00000
6	-981.8	-984.7	0.6122296	.83158
7	-914.9	-918.5	1.9495727	.02887
8	-870.5	-874.7	1.6837787	.06844
9	-798.7	-803.7	1.1602795	.31235
10	-743.3	-749.2	1.4433013	.14698
11	-694.2	-701.0	0.5349320	.89067
12	-628.4	-636.1	1.0644146	.39167

Once the order of periodic autoregressive model is identified, then periodic variation in the parameters of autoregressive model will be measured. To verify periodic variation an F-test on the residual sum of squares with the null hypothesis of non-periodicity is performed. The test is based on the model given by equation (1). In case of the presence of null hypothesis an autoregressive parameter of order 2 is estimated, whereas in the alternating a periodic autoregressive model of order 2 is computed.

TABLE II
TEST OUTPUT FOR PERIODIC VARIATION IN AR PARAMETERS.

Null Hyp	Alt. Hyp	Model Parameters	F-stat	Prob. Value
AR(2)	PAR(2)	*Flow, SI*	7.08	0.00 ***
		Flow, SI,ST	6.72	1.1e-16***

SI=Seasonal Intercept,
ST= seasonal trend,
DF=Degree of freedom
**** = Significance level at 1%*

The result of the test is summarized in Table II. The test output of periodic variation in AR parameters projected in Table II rejects the non-periodicity at the 1% significance level. The test confirms that periodic autoregressive model is adequate to the Lyari waste flow data rather then an autoregressive model.

The univariate representation of PAR (2) for river monthly waste flow time series $Flow_t$ is:

$$Flow_t = \sum_{s=1}^{12} \phi_{1s} Flow_{t-1} + \sum_{s=1}^{12} \phi_{2s} Flow_{t-2} + \varepsilon_t \qquad (12)$$

and values of estimated coefficients are given in Table III.

TABLE III
SEASONAL ESTIMATED COEFFICIENTS
FOR PAR(2) MODELS

Season (s)	ϕ_{1s}	ϕ_{2s}
1	0.04	0.63
2	0.99	0.01
3	1.14	-0.14
4	1.16	-1.12
5	0.98	0.13
6	1	0
7	0.55	0.45
8	1.14	-0.20
9	-0.08	1.05
10	-0.78	1.36
11	0.97	0.01
12	1.77	-0.70

The test statistics of the equation (12) are

Residual standard error = 0.0451 on 358 degrees of freedom

Multiple R-squared = 0.9994

Adjusted R-squared = 0.9994

F-statistic = 1.71e+04 on 36 and 358 degree of freedom

Prob.-value < 2.2e-16

The representation of PAR (2) process given by (12) can be rewritten as vector autoregressive model as

$$\Phi_0 FLOW_T = \Phi_1 FLOW_{T-1} + \varepsilon_T \qquad (13)$$

where $FLOW_T$ represents the observation of waste flow time series in season *s* of year *T*,
with

$$\Phi_0 = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ -0.986 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0.138 & -1.142 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1.120 & -1.161 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -0.131 & -0.985 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -0.999 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -0.447 & -0.555 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0.199 & -1.137 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & -1.051 & 0.075 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1.360 & 0.777 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -0.007 & -0.972 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.703 & -1.768 & 1 \end{bmatrix} \qquad (14)$$

and

$$\Phi_1 = \begin{bmatrix} 0&0&0&0&0&0&0&0&0&0&0.628&0.041\\ 0&0&0&0&0&0&0&0&0&0&0&0.007\\ 0&0&0&0&0&0&0&0&0&0&0&0\\ 0&0&0&0&0&0&0&0&0&0&0&0\\ 0&0&0&0&0&0&0&0&0&0&0&0\\ 0&0&0&0&0&0&0&0&0&0&0&0\\ 0&0&0&0&0&0&0&0&0&0&0&0\\ 0&0&0&0&0&0&0&0&0&0&0&0\\ 0&0&0&0&0&0&0&0&0&0&0&0\\ 0&0&0&0&0&0&0&0&0&0&0&0\\ 0&0&0&0&0&0&0&0&0&0&0&0\\ 0&0&0&0&0&0&0&0&0&0&0&0 \end{bmatrix} \tag{15}$$

The eigen values i.e. $\Phi_0^{-1}\Phi_1$ is calculated for acquiring the information of possible unit root in the $Flow_t$. The estimated eigen values matrix for the Layri flow data set is

$$\Phi_0^{-1}\Phi_1 = [0.06 \quad 0 \quad 0 \quad 0 \quad 0 \quad 0 \quad 0 \quad 0 \quad 0 \quad 0 \quad 0 \quad 0] \tag{16}$$

The eigen value matrix rejects the presence of long run unit root.

The time-varying impact of accumulation of shocks for each season is calculated by utilizing $\Phi_0^{-1}\Phi_1\Phi_0^{-1}$ whose output is projected in Table IV.

TABLE IV
IMPACT ESTIMATED VALUES FOR SEASON TO SEASON.

Season											
1	2	3	4	5	6	7	8	9	10	11	12
0.059	0.121	0.436	0.337	0.342	0.107	0.525	0.925	-0.502	0.652	0.701	0.041
0.059	0.120	0.435	0.336	0.341	0.107	0.523	0.922	-0.500	0.650	0.703	0.047
0.059	0.121	0.436	0.337	0.342	0.107	0.525	0.925	-0.502	0.652	0.706	0.048
0.003	0.005	0.019	0.015	0.015	0.005	0.023	0.041	-0.022	0.029	0.032	0.003
0.010	0.021	0.077	0.059	0.060	0.019	0.092	0.162	-0.088	0.114	0.125	0.009
0.010	0.021	0.076	0.059	0.060	0.019	0.092	0.162	-0.088	0.114	0.124	0.009
0.010	0.021	0.077	0.059	0.060	0.019	0.092	0.162	-0.088	0.115	0.125	0.009
0.010	0.020	0.072	0.056	0.056	0.018	0.087	0.153	-0.083	0.108	0.117	0.009
0.010	0.021	0.075	0.058	0.059	0.018	0.090	0.159	-0.086	0.112	0.122	0.009
0.005	0.011	0.040	0.031	0.031	0.010	0.048	0.084	-0.045	0.059	0.064	0.005
0.005	0.011	0.039	0.030	0.031	0.010	0.047	0.083	-0.045	0.058	0.063	0.005
0.006	0.011	0.041	0.032	0.032	0.010	0.049	0.087	-0.047	0.061	0.067	0.005

.

The impact estimate envisages that season eight has long run impact. i.e. month of August has long run time varying. impact; hence, it is more sensitive to change in stochastic trend

VI. ADEQUACY OF PAR (P) MODEL

The analysis of residual analysis is performed to verify goodness of fit of proposed model.

(i) An F-statistics based test is perform for checking seasonal heteroskedasticiy in the residual of the PAR (2) estimate. This test is asymptotically F distributed with *(11, n-k)* degree of freedom in which n and k denotes, number of observation and number of parameters respectively. The output of the test is

F-statistic = 9.7 on 11 and 393 DF

Prob- value = 1.44329e-15

This test rejects the seasonal heteroskedasticiy at 1% significance level.

(ii) The Ljung-Box test confirms the overall randomness in the residual of proposed model. The test statistics are computed by using (17)

$$Q(\xi) = n(n-2)\sum_{k=1}^{m}(n-k)^{-1}\hat{\xi}_k^2 \tag{17}$$

in which

n = Size of the sample

$\hat{\xi}_k$ = Sample autocorrelation at *lag k*

m = Number of tested lags.

The randomness is rejected if

$$Q(\xi) > \chi^2_{1-\lambda,k} \tag{18}$$

where $\chi^2_{1-\lambda,k}$ is critical region of rejection of the randomness hypothesis at signifivance level λ with k degrees of freedom. The generated output for the Ljung-Box test is as follows:

Ljung test statistic = 29.8498

Degree of freedom = 25

Prob. Value = 0.049

The above computed values indicates that residuals are random at 5% level which signify the appropriateness of PAR (2) model.

(iii) The sketch of the residual shows (see Fig 2) the randomness which implies competence of constructed model.

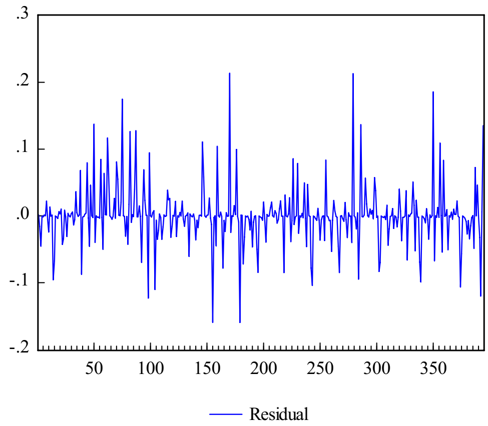

Fig. 2 Residual plot confirms adequacy of the model

(vi) The normality in residual is verified by applying Doornik-Hansen, Shapiro-Wilk, Lilliefors and Jarque-Bera tests. These estimates with their probabilities values in parenthesis yields

Doornik-Hansen test = 120.188 (7.97097e-027)
Shapiro-Wilk test = 0.805041 (1.89447e-021)
Lilliefors test = 0.228565 (~ = 0.00000)
Jarque-Bera test = 641.719 (4.49248e-140)

The above estimated test statistic affirmed that residual series is normally distributed.

(v) The CUSUM test is performed for checking stability in model parameters. The test confirms that PAR (2) parameter are stable as its cumulative sum goes inside the area between the two critical lines. The plot of standardized cumulative recursive residual is illustrated in Fig 3

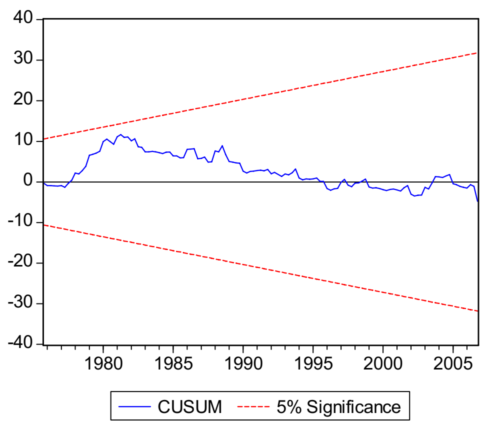

Fig. 3 The CUSUM test shows parameter stability

Moreover, CUSUMSQ (CUSUM of squares) statistic is also within the 5 percent critical lines which give confirmation of parameter/ variance stability exposed by Fig. 4

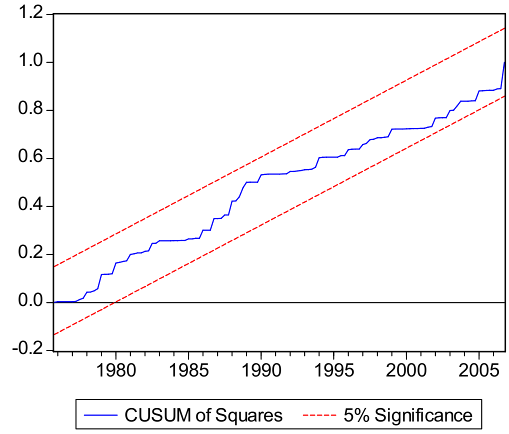

Fig. 4 CUSUMSQ statistics verify parameter stability

In view of the above carried test/ analysis periodic autoregressive model of order two PAR (2) appears to be the most adequate for Lyari waste flow data set.

The out of sample forecast estimated values by using the proposed model are reported in Table V.

TABLE V
PREDICTABLE OUT OF SAMPLE RIVER WASTE FLOW FORECAST

Forecast	Forecast Standard Error	Upper Confidence Interval	Lower Confidence Interval
5.9771	0.0148	5.9169	5.9923
6.0820	0.0176	6.2500	5.5700
6.2134	0.0249	7.0000	5.7100
6.0444	0.0275	6.8300	5.5700
6.6974	0.0311	7.0400	5.8200
6.7452	0.0362	7.4300	6.1500
6.8112	0.0413	7.4300	6.0400
6.8004	0.0558	7.4000	6.0300
6.9757	0.0588	7.6800	6.0000
6.1212	0.0614	7.1100	5.5000
6.2262	0.0647	7.2000	5.6700
6.6426	0.0684	7.5100	5.8000

The graph of the forecasted waste flow along with upper and lower confidence limit is displayed in Fig. 5.

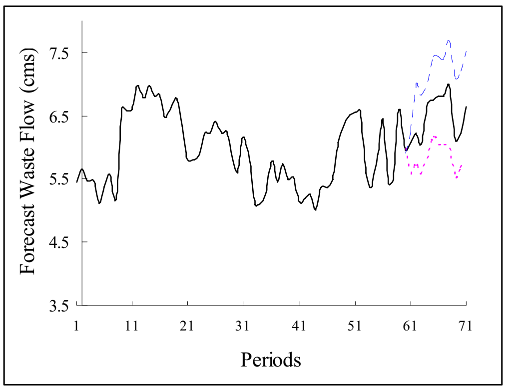

Fig. 5 Time plot of river Lyari waste flow with forecast and 95% confidence limit

The forecast statistics of PAR (2) model are

Root Mean Square Error = 0.0370884

Mean Absolute Error = 0.227902

Mean Absolute Percentage Error = 3.733013

Theil Inequality Coefficient = 0.030929

Bias Proportion = 0.005160

Variance Proportion = 0.116571

Covariance Proportion = 0.878270

The forecasted estimates under Table V, exhibited forecasted plot and test statistics also strengthen the adequacy of PAR (2) model for river Layri waste flow data set.

VII. Conclusion

In the present study PAR model of order 2 is presented for studying monthly waste flow of Lyari river for the available data from January 1974 to December 2006. The study established that proposed model is excellent in assessing periodicities in waste flow of Lyari river. The diagnostic tests confirm adequacy of the proposed model. Moreover, forecast estimate by utilizing the projected model also verified the validity of the model. This model could be of great help for the planning personal and administrative authorities to control the environmental condition of the river, coastal area and to save marine aquatic species.

References

[1] J.D. Salas, G.Q. Tabios III and P. Bartolini, "Approaches to multivariate modeling of water resources time series," *Water Resources Bulletin*, vol. 21, pp. 683-708,1985.

[2] J.D. Salas, *Analysis and modeling of hydrologic time series*. D.R. Maidment, ed., McGraw-Hill, Handbook of Hydrology, 1993, Scetion.19.5-19.9.

[3] G.E.P. Box, G.M. Jenkins, and G. C. Reinsel, *Time Series Analysis: Forecasting and Control*. India: Pearson Education, 2003.

[4] A.C. Harvey, *Forecasting Structural Time Series and Kalman Filter*. Cambridge: Cambridge University Press, 1989.

[5] S. Hylleberg, R.F. Engle, C.W.J. Granger, and B. Yoo, "Seasonal integration and cointegration," *Journal of Econometrics*, vol 44, pp. 215-238,1990.

[6] P.H. Franses, and R. Paap, *Periodic Time Series Models*. United Kingdom: Oxford University Press, 2004.

[7] H.L. Hurd, and N.L. Gerr, "Graphical method for determining the presence of periodic autocorrelation," *Journal of Time Series Analysis*, vol.12, pp. 337-350, 1991.

[8] P. Bloomfield, H.L. Hurd, and R.B. Lund, "Periodic correlation in stratospheric ozone data," *Journal of Time Series Analysis*, vol.15, pp.127-150,1994.

[9] M. Bentarzi, and M. Hallins, "Locally optimal test against periodical autoregression: Parametric and nonparametric approaches," *Econometric Theory*, vol.12, pp. 88-112, 1996.

[10] H. Akaike, "Fitting autoregressive model models for prediction," *Annals of the Institute of Statistical Mathematics*, vol. 21, pp. 243-247, 1969.

[11] G. Schwarz, "Estimating the dimension of a model," *Annals of Statistics*, vol.6, pp. 461-464,1978.