FORECASTING THE POPULATION OF PAKISTAN USING ARIMA MODELS

Muhammad Zakria* and Faqir Muhammad
Department of Mathematics and Statistics, Allama Iqbal Open University, Islamabad
*Corresponding author's e-mail: zakriauaf@yahoo.com

Today, the tremendous increase in population is the major issue of the world. In this empirical study, population of Pakistan from 1951 to 2007 is modelled using Box Jenkins ARIMA methodology. The population of Pakistan is also forecasted for the next 20 years using the parsimonious ARIMA (1, 2, 0) model. If the current growth rate trend continues, the population of Pakistan would be approximately 230.7 million in 2027. The model was validated by the criteria of MSE, AIC, SC, P-values and graphical techniques e.g. ACF and PACF plots of residuals. The projected population by parsimonious ARIMA model is close to the projected population of Pakistan by different bureaus. These bureaus reported 229 million population of Pakistan for the year 2025.
Keywords: Population Census, ARIMA, STATE SPACE, AIC, population forecasting

INTRODUCTION

Now-a-days, the major issue of the world is overpopulation especially of the developing countries. It is unanimously accepted that with the increase of population, the number of associated problems like food, accommodation, education, medical, traffic etc. are generated. Furthermore, the crime rate among the societies also arises due to heavy pressure of the population. Different measures and strategies are being adopted by multicultural societies of the world to limit the size of population according to their feasibility and circumstances. The brief history of the population size of Pakistan during the last 51 years from 1947 to 1998 was 33.74 million, 42.98 million with growth rate 2.45%, 65.31 million with growth rate 3.67%, 84.25 million with growth rate 3.06% and 132.35 million with growth rate 2.69% for the census of 1951, 1961, 1972, 1981 and 1998 respectively (Anonymous, 1967, 1972, 1984, 2001). It indicates that population of Pakistan has increased quadruplet from 1947 to 1998. It is also noted the increasing trend in growth rate during 1951-1972 and decreasing trend from 1972 to 1998. "According to the (Iqbal, 2007), the population of Pakistan was 158.28 million and growth rate 1.98%". Similarly the growth rate was 2.69% according to the 1998 population census. It is evident from the above mentioned figures that growth rate is being decreased as compared to the 1998 growth rates. National Institute of Population Studies Islamabad (2006) also reported the population of Pakistan 156.26 million and growth rate 1.86%.

In Pakistan, the first four population censuses were conducted timely after every ten years with a little bit variation whereas the last census is delayed about seven years due to some national political issues. It is a universal truth that most important asset of any country is the census data as the current and future economy as well as planning of the country depends on census data. The Governments and industrialists of the world also emphasize the significance of population statistics that is inevitable for future planning. Each and every product of the industry, even the wastage of the industry is used by the population in one or the other way. Furthermore all products are designed for national and international population. It might be medical items, food items, gas, patrol, gold etc. In addition the employments of population in different departments like educational institutions, industrial units, stock market and infrastructure depends upon the age and sex distribution of that country. It shows that age and sex distribution of the population is directly related to the consumption of industrial products and plays a significant role in improving a GNP of the country. So the population data is of great importance for national and international Governments, Non Governments Organizations and industrialists etc.

Developed countries are much more aware about their past, current and future population trends, sizes as well as needs. On the other hands, developing countries are least aware about their population and needs. Only those countries are in good economic and social status which is more aware about the current and future population distribution. It is fact that without the optimum knowledge of the population, no country can be on the right track of scientific and technological development in today's world. Keeping in view the importance of population distribution, population census is conducted after every 10 years in most of the countries but some surveys are also conducted after every five years to update population data.

Population data being a most important asset in the present era, the scientists focus on determining the significant trends of the population distribution. That is why; mostly publications are on population trends, size, age and sex distribution. To forecast the population, different linear, nonlinear, first and higher degree regression models, simple and double exponential, logistic regression, simple decay and growth models are being used (Shryock *et al.*, 1973, Jan *et al.*, 2007; Agrawal, 2000). Component method of population projection is the most widely used method which uses the data on fertility, mortality and migration (Srinivasan, 1998). Autoregressive integrated moving average (ARIMA) model is also used to model the time series data in different disciplines.

The main focus of this paper is to forecast the population trend of Pakistan for the next 20 years on the basis of previous trends using a stochastic ARIMA model. Minitab Statistical Software-14 and Eveiw-5 were used for model fitting and forecasting.

METERIALS AND MEHTODS

The data used in this study consist of Pakistan population over the past 57 years (1951-2007). Most of the data is taken from (Kamal *et al.*, 2003) and some of the data is from (Iqbal, 2007).

The ARIMA model technique is not common to forecast the population. Verbeek (2005) give the following general form of ARMA (p, q) model:

$$Y_t = \delta + \phi_1 Y_{t-1} + \ldots\ldots\ldots\ldots + \phi_p Y_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \ldots\ldots\ldots\ldots + \theta_q \varepsilon_{t-q}$$

The major steps involved are: model identification, fitting, validation and forecasting. The procedure adopted for model selection in this study is: to check the stationarity of the series as well as for model identification different time series plots, ACF and PACF are constructed using actual, differenced and transformed data. To transform the population data, Box Cox transformation is used (Box and Jenkins, 1976). After the identification of the model, different stochastic ARIMA models are fitted on the 2nd differenced of logarithmic population series and for the selection of the parsimonious model, different model validation statistics are recommended e.g. Mean squared error (MSE), AIC and P-values etc. These measures are computed for each candidate model and the model having smallest AIC is recommended by assuming to be closest to the unknown reality by which the series is generated (Burnham *et al.*, 2002). Similarly graphical validation approaches are also applied e.g. normal probability plot, histogram, residual plots, and PACF, ACF plots of residuals for the selection of a parsimonious model. After selecting a parsimonious validated model, population of Pakistan is forecasted for the next 20 years along with confidence interval. Both the forecasted and fitted population is plotted on the same plot for model adequacy and comparison purposes in Figure 15.

RESULTS AND DISCUSSION

The time series plot of original population of Pakistan from 1951 to 2007 is given in Figure 1. It exhibits upward increasing trend and suggests that the given time series is non stationary. Figures 2 and 3 depict the trend of the population after taking the 1st and 2nd difference of the original population from 1951 to 2007 respectively. The time series plot in Figure 2 is quite unusual whereas the plot in Figure 3 indicates approximate stationary. Table 1 gives the values of sample ACF, sample PACF, student's t statistics, modified Box-Pierce (Ljung Box) Chi-square statistics, and P-values corresponding to the different lags ranging from 1 to 14. ACF of the population based on Table 1 declines very slowly from high correlation to low correlation during the years 1951-2003. ACF value at lag 1 is 0.95 which is very high as compare to the value of ACF at lag 14 i.e. 0.29 which is not too low. The correlogram of the sample ACF of the original series in Figure 4 indicates high positive correlations at lower lags and low positive correlations at higher lags; furthermore it has an exponential decay which is an indication that the given population series is non stationary series. The most striking feature of this correlogram is that the autocorrelation coefficients at different lags are very high and out of the confidence limits.

One of the major advantages of the correlogram is that it helps in determining the parsimonious model in time series data. To proceed further on the basis of Figure 3 and to know the p, q values of the ARIMA model, the correlogram of the sample partial autocorrelation function (PACF) in Figure 5 has only one spike at lag 1 out of the limits which clearly suggests that an Autoregressive Stochastic model of parameter one i.e. AR (1) seems to be suitable for said series. Box Cox transformation gave the values of λ = 0.22 and its interval (-0.41, 0.85) which contains the value zero (Chatfield, 1996). This recommended that the log transformation is appropriate choice to make our series stationary before to take the difference of the series. Figure 6 shows the trend of the population of Pakistan from 1951 to 2007 using the natural logarithm transformation. The behaviour of the population slightly differs from the trend given in Figure 1.

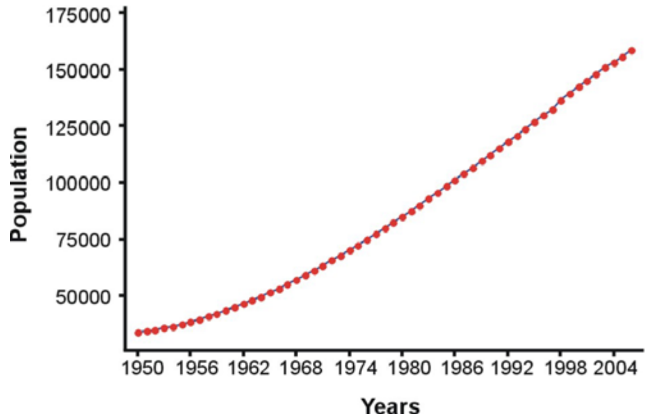

Fig. 1. Population trend of Pakistan during 1951-2007

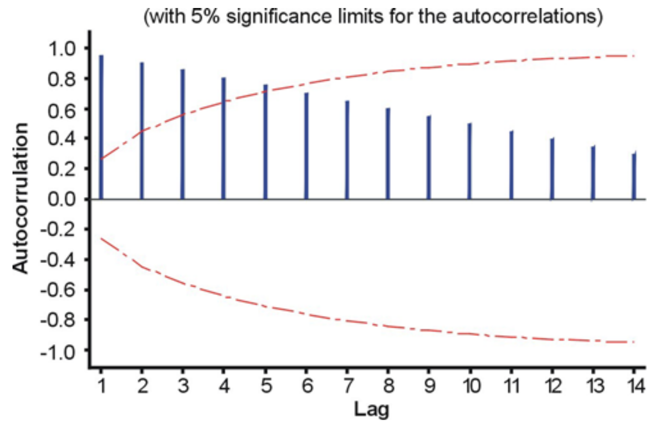

Fig. 4. ACF of Pakistan population during 1951-2007

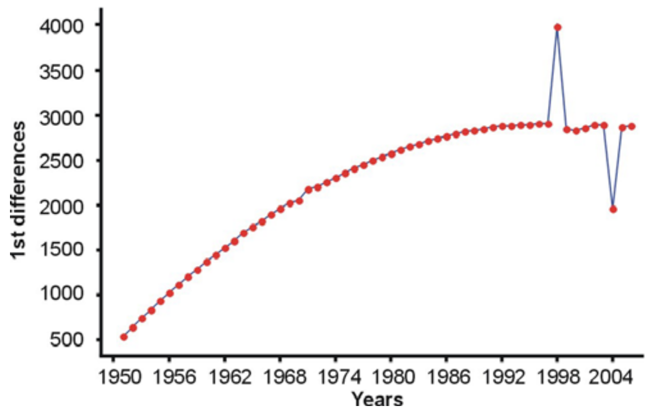

Fig. 2. Population trend of Pak. after 1st differencing (1951-2007)

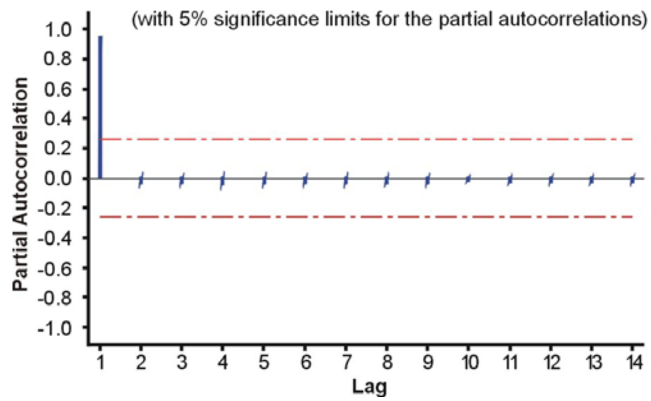

Fig. 5. PACF of Pakistan population during 1951-2007

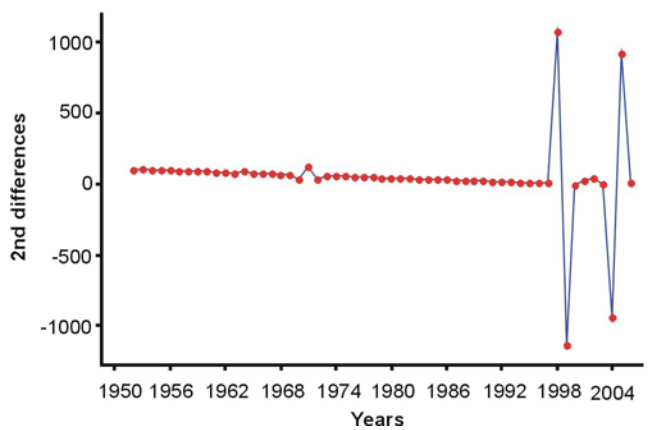

Fig. 3. Population trend of Pak. after 2nd differencing (1951-2007)

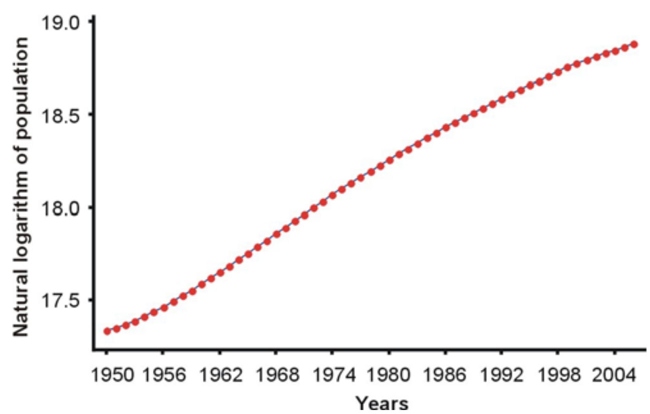

Fig. 6. Loge (Population) trend of Pak. during 1951-2007

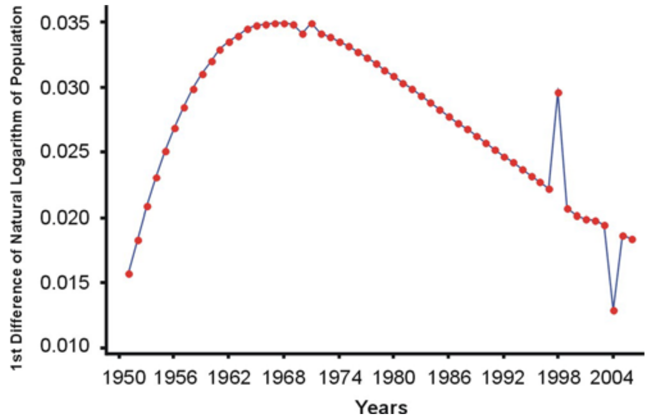

Fig. 7. Loge (Pop.) trend of Pak. After 1st differencing (1951-2007)

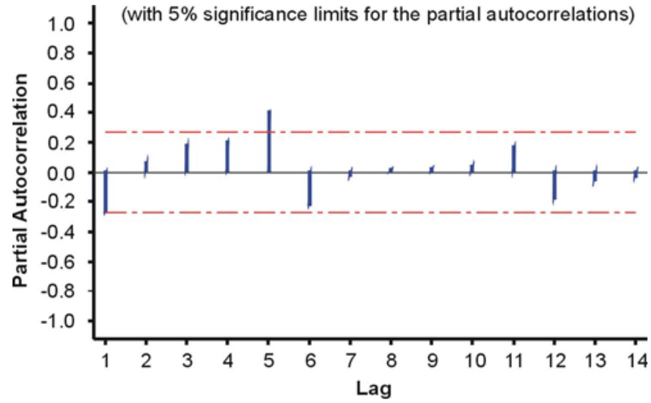

Fig. 10. PACF of Loge (Pop.) of Pak. after 2nd differencing

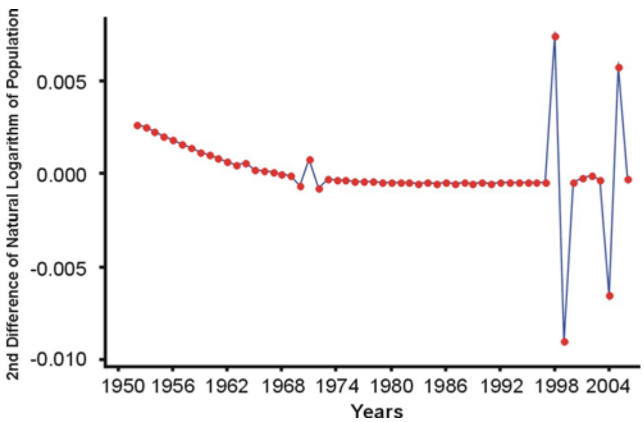

Fig. 8. Loge (Pop.) trend of Pak. after 2nd differencing (1951-2007)

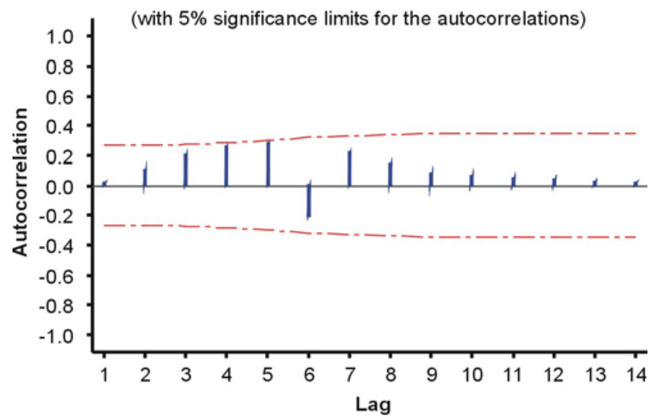

Fig. 11. ACF of residuals using ARIMA (1, 2, 0) on Loge (Population of Pakistan)

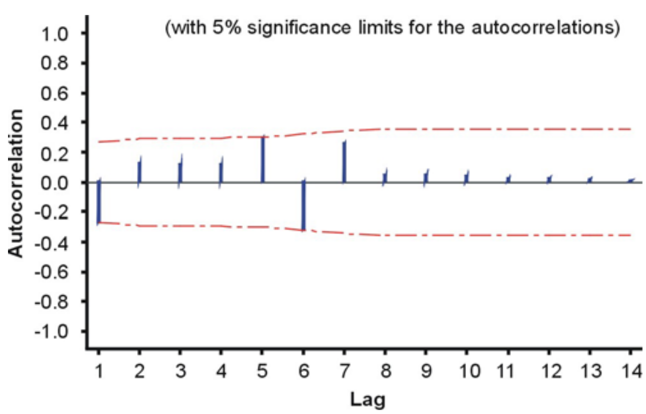

Fig. 9: ACF of Loge (Pop.) of Pak. after 2nd differencing

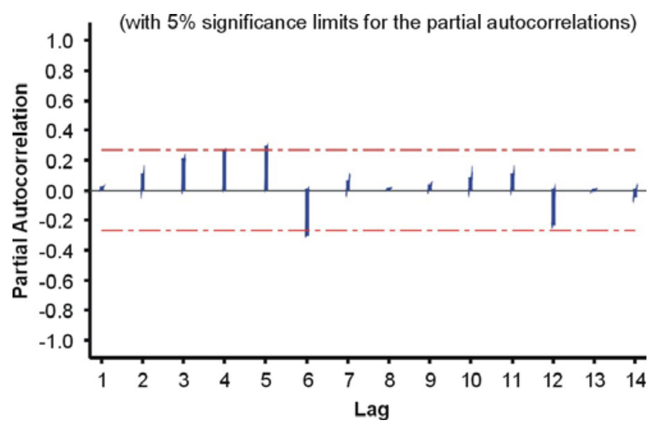

Fig. 12. PACF of residuals using ARIMA (1, 2, 0) on Loge (Population of Pakistan)

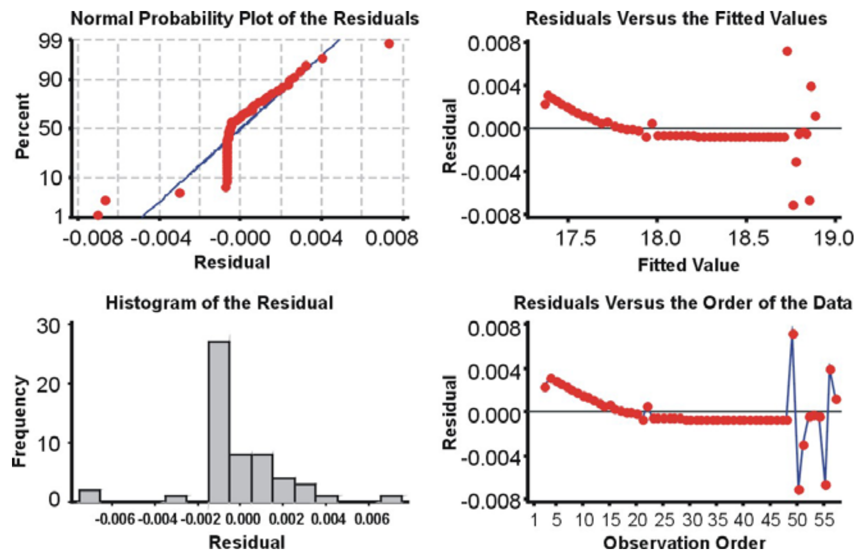

Fig. 13. Four residual plots after 2nd difference of Loge (Population of Pakistan) using ARIMA (1, 2, 0)

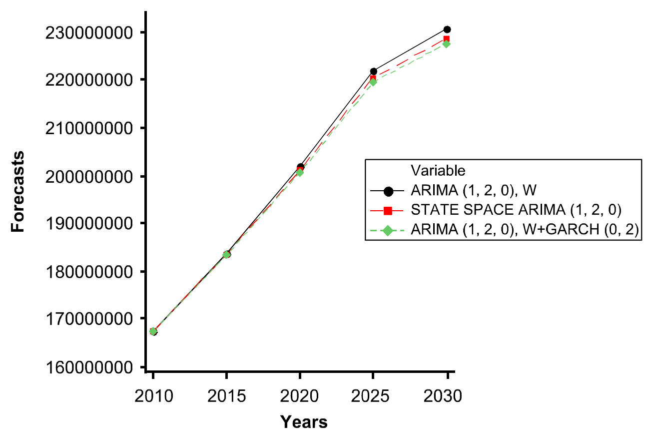

Fig. 14. Time series plot using ARIMA (1, 2, 0), state space ARIMA (1, 2, 0) & GARCH (0, 2)

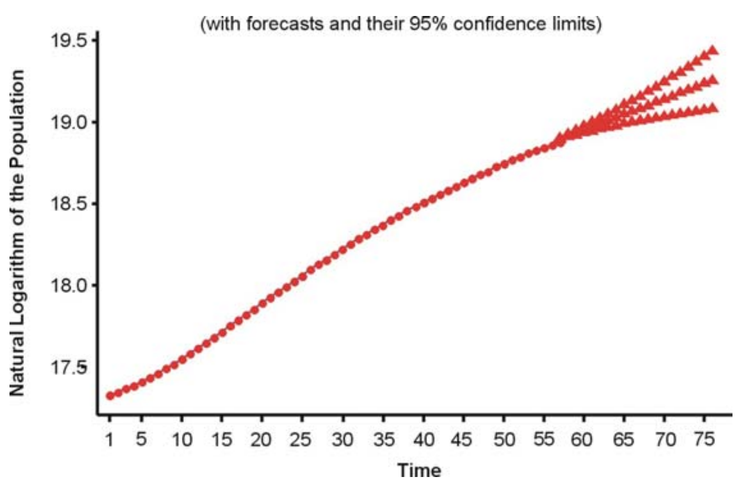

Fig. 15. Logarithmic trend of predicted and projected population using ARIMA (1, 2, 0)

Table 1. Autocorrelation and partial autocorrelation of the actual population of Pakistan

LAG	ACF	PACF	T-STAT	LB-STAT	P-VALUE
1	0.95	0.95	7.19	54.41	0.00
2	0.90	-0.03	4.06	104.26	0.00
3	0.85	-0.03	3.06	149.62	0.00
4	0.80	-0.04	2.50	190.48	0.00
5	0.75	-0.03	2.12	226.98	0.00
6	0.70	-0.03	1.83	259.24	0.00
7	0.65	-0.03	1.60	287.46	0.00
8	0.60	-0.03	1.41	311.82	0.00
9	0.54	-0.03	1.25	332.53	0.00
10	0.49	-0.02	1.10	349.92	0.00
11	0.44	-0.03	0.97	364.22	0.00
12	0.39	-0.03	0.84	375.72	0.00
13	0.34	-0.03	0.73	384.70	0.00
14	0.29	-0.03	0.62	391.46	0.00

Table 2. Autocorrelation and partial autocorrelation of 2nd difference of loge (Population of Pakistan)

LAG	ACF	PACF	T-STAT	LB-STAT	P-VALUE
1	-0.27	-0.27	-1.98	4.13	0.99
2	0.13	0.06	0.90	5.12	0.98
3	0.12	0.18	0.81	5.96	0.97
4	0.12	0.21	0.81	6.84	0.94
5	0.30	0.41	1.99	12.46	0.57
6	-0.32	-0.22	-1.98	18.88	0.17
7	0.26	-0.03	1.52	23.29	0.06
8	0.05	0.02	0.27	23.44	0.05
9	0.05	0.03	0.27	23.60	0.05
10	0.04	0.04	0.23	23.71	0.05
11	0.03	0.18	0.17	23.78	0.05
12	0.03	-0.18	0.14	23.82	0.05
13	0.02	-0.06	0.09	23.84	0.05
14	0.01	-0.03	0.06	23.85	0.05

Figure 7 and 8 present the trend of the population after taking the first and second difference of the natural logarithm of the population series during the same period. Figure 7 is unusual as that of Figure 2. Figure 8 is approximate stationary and has more sophisticated trend than Figure 3. The 2nd difference of natural logarithm of the population series is being used for onward analysis.

Figure 9 is the sample autocorrelation function (ACF) of the 2nd difference of natural logarithm of the population of Pakistan. Almost all the spikes at different lags are within the 95% confidence limits; this is an indication that the selected parsimonious model might be without moving average components. Figure 10 is the sample partial autocorrelation function (PACF) of the same series used in Figure 9. All its spikes in Figure 10 at different lags are within the 95% confidence limits except two spikes, one at lag 1 and second at lag 5. The spike at lag 5 is clear cut out of the positive limit whereas spike at lag 1 is close to the negative limit; other spikes at different lags in PACFs are clearly within the 95 % limits. The values of sample ACF; sample PACF, Students t statistics, Ljung Box statistics, and P-values corresponding to the different lags from 1 to 14 of 2nd difference of natural logarithmic population series are given in Table 2. The characteristics of this correlogram are entirely different as that of the characteristics of correlogram of the actual population in Figure 4 and 5. It is concluded that all the autocorrelation coefficients (ACF) and partial autocorrelation coefficients (PACF) given in Table 2 do not differ significantly from zero and consequently the 2nd difference of the natural logarithm of the population series seems to be stationary. This indicates that different stochastic stationary models can be studied on this series.

Table 3 consists on parameter estimates and goodness of fit measures of fourteen ARIMA models. Last three rows of Table 3 present different models but these models provide the same information. According to the scientific approach, if more than one model provides same information, the researcher should recommend the insignificant model which has minimum number of parameters for ease estimation and simple interpretation of parameters. If (P-value ≤ 0.05) corresponding to an estimate in the model, the hypothesis that the parameter equal to zero is rejected on the other hand If (P-value ≥ 0.05) corresponding to an estimate in the model, the hypothesis that the parameter equal to zero is not rejected which suggests that the explanatory variable should not be included in the model. Out of three models of Table 3, only one parsimonious model is to be selected i.e. ARIMA (1, 2, 0). Actually in time series studies, the selection of the parsimonious model is an art not a mathematical science. In real phenomena, it is tried to model the real situation that is why, level of significance has to be relaxed to fulfil the assumptions. So the recommended parsimonious model is AR (1).

Figure 11 and 12 are the ACF and PACF of the residuals using the model ARIMA (1, 2, 0) on the natural logarithm of the population of Pakistan. All the autocorrelation and partial autocorrelations of the residuals at different lags are within the 95 % confidence limits which supports and strengths the recommended parsimonious stochastic time series model. Also there is an indication about the adequacy of the above said stochastic time series models.

Figure 13 presents the different graphical measures for the adequacy of the model. The first measure is the normal probability plot of the residuals which is not as

Table 3. Parameter estimates and goodness of fit of different ARIMA (p, d, q) models

ARIMA (p,d,q)	$Y_t = \delta + \phi_1 Y_{t-1} + \ldots\ldots\ldots + \phi_p Y_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \ldots\ldots\ldots + \theta_q \varepsilon_{t-q}$							
ARIMA (5,2,1),W	ϕ_1	ϕ_2	ϕ_3	ϕ_4	ϕ_5	θ_1	θ_2	Intercept
Coefficients (Coeff.s P-value)	-0.633 (0.001)	-0.138 (0.324)	0.331 (0.018)	0.667 (0.000)	0.801 (0.000)	-0.243 (0.291)		-6.87558E-05 (0.919)
MSE 0.0000028, AIC -9.908, BIC -9.641, D.W. Stat. 1.951, P-value 0.391								
ARIMA (5,2,1) (Coeff.s P-value)	-0.732 (0.001)	-0.181 (0.250)	0.192 (0.214)	0.471 (0.006)	0.654 (0.000)	-0.421 (0.104)		
MSE 0.0000032, AIC -9.898, BIC -9.669, D.W. Stat. 1.940, P-value 0.844								
ARIMA (2,2,2), W (Coeff.s P-value)	1.174 (0.000)	-0.166 (0.328)				1.731 (0.000)	-0.866 (0.000)	-2.28224E-05 (0.764)
MSE 0.0000034, AIC -10.164, BIC -9.978, D.W. Stat. 2.208, P-value 0.077								
ARIMA (2,2,2) (Coeff.s P-value)	0.699 (0.004)	-0.317 (0.160)				-0.317 (0.000)	-0.812 (0.000)	
MSE 0.0000039, AIC -9.793, BIC -9.64, D.W. Stat. 2.034, P-value 0.331								
ARIMA (2,2,0), W (Coeff.s P-value)	-0.253 (0.0740)	0.068 (0.653)						0.000 (0.834)
MSE 0.0000045, AIC -9.458, BIC -9.347, D.W. Stat. 2.058, P-value 0.025								
ARIMA (2,2,0) (Coeff.s P-value)	-0.253 (0.0710)	0.065 (0.6620)						
MSE 0.0000044, AIC -9.495, BIC -9.421, D.W. Stat. 2.058, P-value 0.039								
ARIMA (1,2,1), W (Coeff.s P-value)	1.002 (0.000)					1.047 (0.000)		-7.45839E-07 (0.977)
MSE 0.0000047, AIC -9.531, BIC 8.669, D.W. Stat. 2.866, P-value 0.001								
ARIMA (1,2,1) (Coeff.s P-value)	-0.360 (0.487)					-0.093 (0.865)		
MSE 0.0000044, AIC -9.522, BIC -9.448, D.W. Stat. 2.698, P-value 0.028								
ARIMA (1,2,0) (Coeff.s P-value)	-0.273 (0.042)							
MSE 0.0000044, AIC -9.501, BIC -9.464, D.W. Stat. 1.982, P-value 0.034								
ARIMA(0,2,1), W (Coeff. P-value)						0.213 (0.120)		0.000 (0.882)
MSE 0.0000044, AIC -9.442, BIC -9.369, D.W. Stat. 2.032, P-value 0.016								
ARIMA (0,2,1) (Coeff. P-value)						0.213 (0.115)		
MSE 0.0000045, AIC -9.478, BIC -9.441, D.W. Stat. 2.029, P-value 0.026								
ARIMA (1,2,0), W (Coeff.s P-value)	-0.270 (0.046)	$Y_t^* = \phi_0 + \phi_1 Y_{t-1}^* + \varepsilon_t$						0.000 (0.954)
MSE 0.0000044, AIC -9.464, SC -9.390, D.W. Stat. 1.981, P-value 0.022								
ARIMA (1,2,0) State Space (Coeff.s P-value)	$Y_t^* = \varphi + \phi_1 Y_{t-1}^* + \varepsilon_t$							
	φ			ϕ_1				
	-12.367 (000)			-0.272 (0.000)				
	MSE 0.0000043, AIC -9.455, SC -9.382, HQ -9.427							
ARIMA (1,2,0) GARCH (0,2) (Coeff.s P-value)	$Y_t^* = \phi_0 + \phi_1 Y_{t-1}^* + \gamma_0 + \omega_1 \sigma_{t-1}^2 + \omega_1 \sigma_{t-2}^2$							
	ϕ_0	ϕ_1		γ_0		ω_1		ω_1
	-2.19E-05 (0.907)	-0.389 (0.000)		1.92E-07 (0.000)		1.868 (0.000)		-0.923 (0.000)
	AIC -9.567, SC -9.383, D.W. Stat. 1.718							

Table 4. Projected population for the years 2010, 2015, 2020, 2025, 2027 using different models

Years Model	2010 (million)	2015 (million)	2020 (million)	2025 (million)	2027 (million)
ARIMA (5,2,1),W	165.49	177.78	188.22	196.29	198.90
ARIMA (5,2,1)	166.09	180.57	195.76	211.91	218.69
ARIMA (2,2,2),W	160.72	165.40	171.44	171.98	234.31
ARIMA (2,2,2)	167.57	184.98	204.13	225.26	234.31
ARIMA (2,2,0),W	167.48	184.20	184.20	223.66	232.66
ARIMA (2,2,0)	167.42	183.84	201.86	221.65	230.10
ARIMA (1,2,1),W	165.99	175.26	179.29	177.65	175.40
ARIMA (1,2,1)	167.33	183.56	201.37	220.91	229.24
ARIMA (0,2,1),W	167.20	183.33	201.19	220.97	229.47
ARIMA (0,2,1)	167.16	183.10	200.56	219.67	227.81
ARIMA (1,2,0)	167.27	183.39	201.07	220.46	228.72
ARIMA (1,2,0),W	167.31	183.67	201.82	221.99	230.68
STATE SPACE ARIMA (1,2,0)	167.27	183.40	201.08	220.46	228.73
ARIMA(1,2,0),W + GARCH (0,2)	167.24	183.23	200.64	219.59	227.62

*W indicates the inclusion of intercept in the model.

Table 5. Projected population using ARIMA (1, 2, 0) with intercept during 2008-2027

Year	Natural log of projected population	Projected population (million)	Lower limit of projected population (million)	Upper limit of projected population (million)
2008	18.90	161.23	160.56	161.90
2009	18.92	164.24	162.89	165.60
2010	18.94	167.31	165.10	169.55
2011	18.95	170.45	167.23	173.73
2012	18.97	173.65	169.28	178.36
2013	18.99	176.92	171.26	182.77
2014	19.01	180.26	173.18	187.63
2015	19.03	183.67	175.04	192.72
2016	19.05	187.15	176.84	198.10
2017	19.07	190.70	178.58	203.64
2018	19.09	194.33	180.28	209.48
2019	19.10	198.04	181.92	215.60
2020	19.12	201.82	183.51	221.96
2021	19.14	205.69	185.05	228.63
2022	19.16	209.64	186.55	235.58
2023	19.18	213.67	188.10	242.85
2024	19.20	217.79	189.40	250.43
2025	19.22	222.00	190.75	258.35
2026	19.24	226.29	192.06	266.62
2027	19.26	230.68	193.33	275.25

good as required for an adequate model although some of the residuals in this plot are in scattered positions but most of the residuals are on the straight line. Second measure for adequacy of model is the histogram of the residuals which does not show exact normality of the residuals but luckily the majority of the residuals are at centre. Some of the residuals are very large which complicate the situation or the main source of difficulty. Third and fourth measures are the plot of residuals Vs fitted values and order of the data respectively. Almost all of the residuals are within acceptable limits which indicate the adequacy of the recommended model.

Table 4 consists on the forecasted population for the next 20 years including the years 2010, 2015, 2020, 2025, and 2027 using different ARIMA models. On the basis of goodness of fit criteria, again the same three models provide the consistent results and out of which AR (1) is the simplest one which strengths the recommended parsimonious model.

Figure 14 presents the comparative trend of last three models given in Table 4 which are ARIMA, State Space ARIMA and GARCH models. Overall the forecast with these three time series models are almost same during the next 20 years but after 10 years in 2017, these models slightly differ with each other. The forecast by GARCH model is slightly less than State Space model whereas forecast with State Space model is little bit less than the recommended ARIMA model.

Table 5 consists of the forecasted population using ARIMA (1, 2, 0) model for the next 20 years. The second column is the forecasted population in the form of natural logarithm whereas third column is the forecasted population (in million) from 2008 to 2027. Fourth and fifth columns are the lower limits and upper limits of the forecasted population during the years 2008-2027. If the current growth rate remains continue, population of Pakistan would be approximately 230.7 million with 95 % confidence limits 193.3 million & 275.24 million in 2027. Figure 15 indicates the trend of the actual and forecasted population along with the confidence interval for the next 20 years.

CONCLUSIONS AND RECOMMENDATIONS

It is concluded that the ARIMA (1, 2, 0) is an appropriate and parsimonious model to forecast the population of Pakistan for the next 20 years. The Akaike information criteria's value for this model is -9.46 which is approximate nearer as that of the other fitted models given in Table 3 but this model is the simplest model. P-value = 0.022 indicates that the residuals are independent and random. The forecasted population using the above mentioned model for next specific years i.e. 2010, 2015, 2020, 2025, and 2027 is given in Table 4. The forecasted population of Pakistan during the year 2027 would be 230 million and 228 million using simple ARIMA (1, 2, 0) and ARIMA (1, 2, 0) with State Space Kalman filter approach respectively. 95% confidence limits of ARIMA (1, 2, 0) are (193.33 million, 275.25 million). Forecasted population using the ARIMA model with state space Kalman filter estimation approach are almost same as that of simple ARIMA. According to the parsimonious model, there will be 74.29% increase in the Population of Pakistan till 2027 as compared to the 1998 population census where as 45.74% increase in Pakistan population as compared to the IMF World Bank 2007 estimates. "NIPS (2006) estimated the population of Pakistan for different years which are 161.86 million, 175.65 million, 189.42 million and 202.11 million during the years 2010, 2015, 2020, 2025 respectively". These estimates are less than the estimates reported by all other national as well as international bureau. "According to the World Population Prospects (2006), the projected population of Pakistan for different years are 173.35 million, 190.66 million, 208.32 million and 224.96 million for the years 2010, 2015, 2020 and 2025 respectively". These estimates are higher than the estimates by ARIMA (1, 2, 0). Approximately 7 million peoples are more in each of the first three estimates and about 3 million people are less in last projected population. "Rauf Textile and Printing Mills (2008) also reported 213 million population of Pakistan by the year 2025".These estimates are lower as compared to the Population reference bureau (2007), United States Census Bureau (2008), Pakistan reality (2008) and Population of Pakistan (2008) on the other hand these estimates are greater than the estimates of (NIPS, 2006). The projected population given in Table 4 is slightly higher than the estimates of World Population Prospects, 2006. These estimates are close to the majority of the other estimates. It might be due to the use of different growth models and methodologies. The ARIMA methodology is simple and straight forward and needs least information. The projected population by ARIMA (1, 2, 0) is close to the projected population of Pakistan by different bureaus i.e. Population reference bureau (2007), United States Census Bureau (2008), Pakistan reality (2008) and Population of Pakistan (2008). These bureaus reported 229 million population of Pakistan for the year 2025. Population Resource Centre (2007) mentioned that the demographic factors particularly the youth bulge are causing to unrest in Pakistan. Pakistan being the world's sixth largest country has a population of 165 million. It is also reported that population would be 228 million by 2025 which is almost same to our projected population. In short, the estimates provided in Table 4 using ARIMA (1,2,0) are close to other researcher's finding and are equally important for Government of Pakistan as well as Non Government Organizations for future planning and projects.

REFERENCES

Agrawal, U.D. 2000. Population projections and their accuracy. B.R. Publication Corporation, Delhi.

Burnham, K.P. and D.R. Anderson. 2002. Model selection and multi model inference: A practical information theoretic approach. Springer Science + Business Media. Inc., New York.

Box, G.E.P. and G.M. Jenkins. 1976. Time series analysis: Forecasting and control. Holdan day, San Francisco.

Chatfield, C. 1996. Simple descriptive techniques: In C. Chatfield & J.V. Zidek (Eds.). The analysis of time series: An introduction, pp.9-26. Chapman and Hall, New York.

Census Organization. 1967. Projections of population of Pakistan 1961 to 1981. (Census Bulletin No.7), Ministry of Home & Kashmir Affairs, Home Affairs Division, Government of Pakistan, Islamabad.

Iqbal, Z. 2007. Pakistan: The housing market thrives as more Pakistan capital returns homes. Dawn. p.1.

Jan, B., I. Ali and S. Shah. 2007. Selecting of mathematical model for projections of NWFP population. The Journal of Humanities and Social Sciences. XV(2): 69-78.

Kamal, A.R., I. Muhammad and N. Mahmood. 2003. Population of Pakistan: An analysis of 1998 population and housing census. Pakistan Institute of Development Economics, Islamabad.

National Institute of Population Studies. 2006. Population growth and its implications. Government of Pakistan, Islamabad.

Population Census Organization. 1972. Statistical report of Pakistan: Population Census of Pakistan-1972. Statistics Division, Government of Pakistan, Islamabad.

Population Census Organization. 1984. 1981 Census Report of Pakistan. Statistics Division, Government of Pakistan, Islamabad.

Population Census Organization. 2001. 1998 Census Report of Pakistan. Statistics Division, Government of Pakistan, Islamabad.

Population Resources Centre providing the demographic dimensions of public policy. 2007. Instability in Pakistan: Demographic Factors. Retrieved September 8, 2008, from http://www.prcdc.org/events/15.

Population Reference Bureau. 2007. Retrieved September 10, 2008, from http://www.prb.org/Datafinder/Geography/Summary.aspx?region=145 &region_type=2.

Pakistan Reality. 2008. Projection of population of Pakistan. Retrieved September 10, 2008, from http://www.pakreality.com/

Population of Pakistan. 2008. Population projection of Pakistan. Retrieved September 12, 2008, from http://www.pakistans.com/pakistan/population/

Rauf Textile and Printing Mills (Pvt) Ltd. 2008. Pakistan facts and figures. Retrieved September 13, 2008, from http://rauf.com/Facts_figures.html

Shryock, H.S. and J.S. Seigel. 1973. The methods and materials of demography. U.S. Bureau of the Census. Government Printing Office, Washington, DC.

Srinivasan, K. 1998. Basic demographic techniques and applications. Sage Publications, Delhi.

United States Census Bureau. 2008. International data base. Retrieved September 16, 2008, from http://www.census.gov/ipc/www/idb/country/pkportal.html

Verbeek, M. 2005. Univariate time series models. In: A guide to modern econometrics. pp.255-308. John Wiley & Sons, England.

World Population Prospects. 2006. Population division of the department of economic and social affairs of the United Nations Secretariat. Retrieved September 20, 2008, from http://esa.un.org/unpp.