ARIMA Model Building and Forecasting on Imports and Exports of Pakistan

Ahmad Farooqi
University of Windsor
farooqi868@uwindsor.ca

Abstract

From the day one, mankind has always been interested into the future. As the civilization advanced with growing sophistication in all phases of life, the need to look into the future also grew with it. Today every government, public private organizations, as well as an individual would like to predict and plan for the future. In order to attain a better growth in the economy of a country, modeling and forecasting is the most important tool now a day, this can be done by one of the statistical technique called a Time series analysis. In this paper we tried to build a time series model called ARIMA (Auto Regressive Integrated Moving Average) model with particular reference of Box and Jenkins approach on annually total Imports and Exports of Pakistan from the year 1947 to the year 2013 with useful statistical software R. Validity of the fitted model is tested using standard statistical techniques. The fitted model is then used to forecast some future values of Imports and export of Pakistan. It is found that an ARIMA (2, 2, 2) and ARIMA (1, 2, 2) model looks suitable to forecast the annual Imports and Exports of Pakistan respectively. We also found an increasing trend both in case of Imports and Exports during this study.

Notation and Abbreviation

The following notation and abbreviation are used.

ACF: Auto-correlation function, PACF: Partial autocorrelation function, AR: Auto regressive, MA: Moving average, ARMA: Auto regressive moving average, ARIMA: Auto regressive integrated moving average, AIC: Akaikae information criteria, PKR: Pakistani Rupees

1. Background and Introduction to Imports and Exports of Pakistan

Mankind has always been interested in the future from the early day's soothsayers and fortune tellers have held a respective place in the society. As civilization advanced with growing sophistication in all phases of life, the need to look into the future grew with it. Today every government, NGOS, business, industry as well as an individual wants to be able to predict and plan for the future. In order to attain better growth in the economy of a country, modeling and forecasting is the most important tool now a day. It has been always a challenge for national trade policy makers to identify the trend in an import/export and to defy the gap between these two. Time series analysis and its applications have become increasingly important in various fields of research, such as business, economics, engineering, medicine, environmentalism, social sciences, politics, and others. Pakistan is a developing country with a population of almost 190 million where the economy is mainly based on agriculture. From 1947 to date, Pakistan has been experiencing a trade deficit. Though, Pakistan has made good progress in both exports and imports but the imports have grown relatively more as compared to the exports. As a result, Pakistan is now facing trade deficit, which has become more severe with the passage of time (Shair Ali, 2006). Pakistan's single largest import category is petroleum and petroleum products, other Imports include: industrial machinery, construction

machinery, trucks, automobiles, computers, computer parts, medicines, pharmaceutical products, food items, civilian aircraft, defense equipment, iron, steel, toys, electronics, and other consumer items. On the other hand Pakistan is a second largest exporter of textiles in the world. Even though most produce is substandard, it fares well due to its cheapness. The export commodities of Pakistan include textiles, rice, leather goods, chemicals, and carpets. On the other hand the import commodities include petroleum and petroleum products, machinery, transportation equipment, edible oils, and tea (Economy Watch, 2010). Pakistan commercial policies are administrated by different government. The best way to characterize Pakistan's economic strategy during the first decade of Pakistan came into being is to say that the policies of the government were restricted to the management of short term crises and reaction to various developments in the foreign trade sectors (Zaidi, 2000). After that an Export Promotion Bureau was formed at an institutional level, this step did seem to bear the desired result and at the end of 1968 Exports was almost tripled the efforts of liberalize economy received a set back as a result of the 1965 war with India due to the cost imposed by war bad harvest and slowing down external assistance. After the separation of East Pakistan, Bhutto government took over Pakistan in 1971 this government was totally different from past in almost every respect. In 1972 the currency was again devaluated by 58% as an oil price increase in the world and the economy of the world goes into recession. By the formation of OPEC in 1973 the price of oil was fourfold increase in just a few weeks and as Pakistan being an oil importer suffered tremendously at the end of 1980. Zia government took over Pakistan in 1977. This period was marked by relatively stable as Zia's polices were almost a reversal of the previous government. During 1977 to 1981 there were sharp increases in Imports of Pakistan due to an upswing of world trade (Zaidi, 2000). During 1980s Soviet invaded Afghanistan and as a result of this Pakistan got foreign aid which allows the policy maker to pursue a relatively easy soft option of growth, while Imports expansion continued unabated. The (1988- 1999) period can be characterized as a time of political consensus on the nature of the country's trade policies. This period was the only in the history of Pakistan, where Exports grew at a faster pace than Imports, other hand growth rate of import was 6% during this period, mean while by the decline in an oil price in early 90s reduced the import bill. Pakistan's economic outlook has brightened in Musharraf's era, in conjunction with rapid economic growth and a dramatic improvement in its foreign exchange position as a result of its current account surplus and a consequent rapid growth in hard currency reserves. Construction activity, sales of durable goods such as trucks and automobiles, and housing purchases have all jumped to record levels. Despite rapid growth in domestic automobile manufacturing, Imports have also risen to meet the increased demand. Pakistan's Exports increased more than 100% from 1999 in the year 2008. From 2009 to 2010 the export target of Pakistan was US $20 billion. As of April 2011, Pakistan's Exports stand at US $25 billion. According to the provisional figures compiled by the Pakistan Bureau of Statistics, Exports from Pakistan during the year 2013 reporting a decrease as compared to the year 2012. On the other hand Imports into Pakistan during the year 2013 reported increased as compared to the year 2012.

2. Theoretical Basis of Time Series and ARIMA Models

Time series analysis and its applications have become increasingly important in various fields of research, such as business, economics, engineering, medicine,

environmentalism, social sciences, politics, and others. In this section we will define some definitions related to study. A model that explains the pattern or variation in an actual time series data is known as a time series model. The term Stationarity is defined as a quality of a process in which the statistical parameters (mean and standard deviation) of the process do not change with time (Challis and Kitney November 1991). An auto-correlation coefficient measures the correlation between successive observations of time series data at lag k denoted by r_k and defined by the equation 2.1

$$r_k = \frac{\sum_{t=1}^{N-k}(Y_t - \bar{Y})(Y_{t+k} - \bar{Y})}{\sum_{t=1}^{N}(Y_t - \bar{Y})^2} \tag{2.1}$$

A correlation between observation Y and $Y_t - K(Y_t + K)$ by removing all other influences of the other lags called partial auto-correlation.

Exponential smoothing methods are useful for making forecasts, and make no assumptions about the correlations between successive values of the time series. However, if we want to make prediction intervals for forecasts made using exponential smoothing methods, the prediction intervals require that the forecast errors are uncorrelated and normally distributed with mean zero and constant variance. In practice, most of the time series are non-stationary in order to fit stationary models like MA, AR, ARMA it is necessary to convert the non-stationary series into stationary by taking some possible transformation called integrated model. Box and Jenkins (1976) put together in a comprehensive way to understand the use of univariate time series ARIMA model. Thus the ARIMA model of order (p, d, q) denoted by ARIMA (p, d, q) and given by the equation 2.2.

$$W_t = \mu + \frac{\theta(B)}{\varphi(B)} a_t \tag{2.2}$$

Where t is indexes time, W_t is the response series Y_t or a difference of a response series, μ is the mean term, B is the backward shift operator that is, $BX_t = X_{t-1}$, φ(B) is the autoregressive operator, represented as a polynomial in the backshift operator $\phi(B) = 1 - \phi_1 B - \ldots - \phi_p B^p$, $\theta(B)$ is the moving average operator, represented as a polynomial in the backshift operator: $\theta(B) = 1 - \theta_1 B - \ldots - \theta_q B^q$ and a_t is the independent disturbance, also called the random error.

2.1 ARIMA Model Building Process

This study focuses on the Box-Jenkins (1976) approach to identification, estimation, diagnostic and forecasting a univariate time series models. Before doing any analysis of time series, one can plot the data by using standard plots and summary statistics to see the behavior of the data. The techniques of model identification which are most commonly used were propounded originally by Box and Jenkins. Their basic tools were the sample ACF and the PACF pattern to see whether given data is stationary in its level and variability, if not some possible transformations can be made, for example by taking the differences of data values we can make data stationary in its level on the other hand variability can be handled by taking log of the values, sometime both differences and log

can be made in ordered to make data stationary, which is not always possible. We have to see the outliers and document these with full explanation. Next the estimation of parameters in ARIMA model can be made with a nonlinear least square method, maximum likelihood or method of moments. It is to be noted that estimating the parameters for the Box-Jenkins models are quite complicated nonlinear estimation problem. For this reason, the parameter estimation should be left with a high quality software program that fits Box-Jenkins models. Fortunately, many commercial statistical software programs now fit Box-Jenkins models. It is very important to remember the principle of parsimony which means trying to fit a model having minimum number of parameters as a simple model always outperform more complex model. A good model should have statistically significant coefficients and low AIC or BIC as compared to the other fitted model. Finally a diagnostic check for fitted models is made to analyze the residuals from the fit for any signs of non–randomness. A Box–Pierce and Ljung–Box (1978) test can be used to examine the Null of independently distributed residuals. It's derived from the idea that the residuals of a "correctly specified" model are independently distributed.

3. Data Analysis

The data about annual total Imports and Exports of Pakistan were collected from Pakistan Bureau of Statistics, Islamabad. The data comprise of yearly basis and were in thousand Pakistani rupees from years 1947 to 2013 and were converted into million Pakistani rupees.

3.1 Summary Statistics for Imports

The summary statistics for Import data of Pakistan are given below in the table-1 by using R with function *describe (data*) from the psych-library.

Table 1: Summary Statistics of Imports of Pakistan

N	Mean	Median	Standard deviation	Min	Max	Skewness	Kurtosis	Standard error
67	53689.65	5354.37	104345.90	31.88	400909.30	2.23	3.81	12891.88

3.2 Model Identification for Imports

It can be seen from the graph by *plot (data)* function in R that, the Imports of Pakistan are slowly increasing and decreasing over time up till the year 2000 and after that they were continually increasing up to the year 2010. R provides the functions *acf (data)* for computing and plotting of ACF. The plot of AFC shows that the sample autocorrelations are very strong, positive and decay slowly, which indicates that there are possible shifts in both the mean and the variability over time for this series, it means the arithmetic mean may be edging upwards, and the variability may be increasing. Mean trend can be removed by differencing once or twice and the variability can be controlled by taking

natural logarithm of the given data which is not always possible, also there is a usual problem of the increased difficulty of interpretation of transformed variables. It is well-known that an instantaneous nonlinear transformation applied to the optimal forecast of a variable may not result in the optimal forecast of the transformed variable (Granger and New bold (1976)). In particular, if optimal forecasts of the logs are available, converting them to forecasts for the original variable by applying the exponential function will in general may not be optimal. Therefore, the untransformed difference is used in future analyses.

Figure-1 Graph of Imports of Pakistan

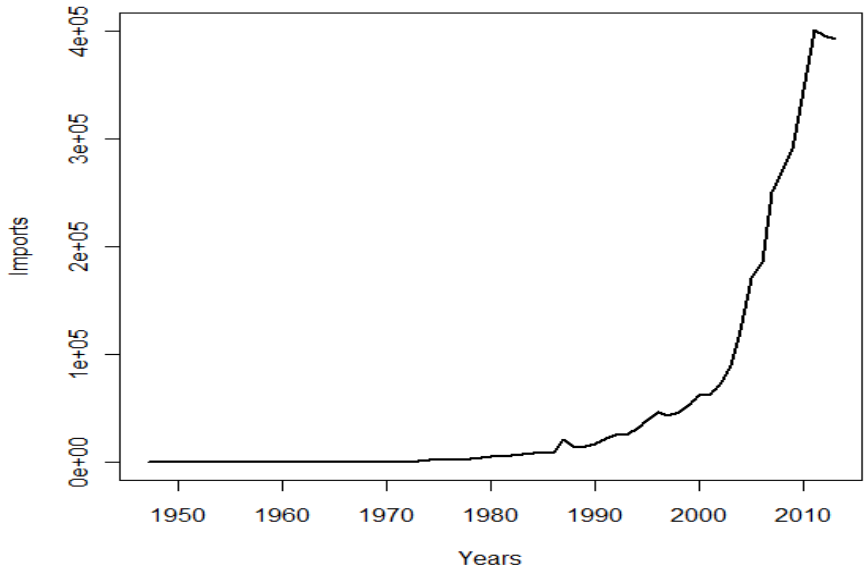
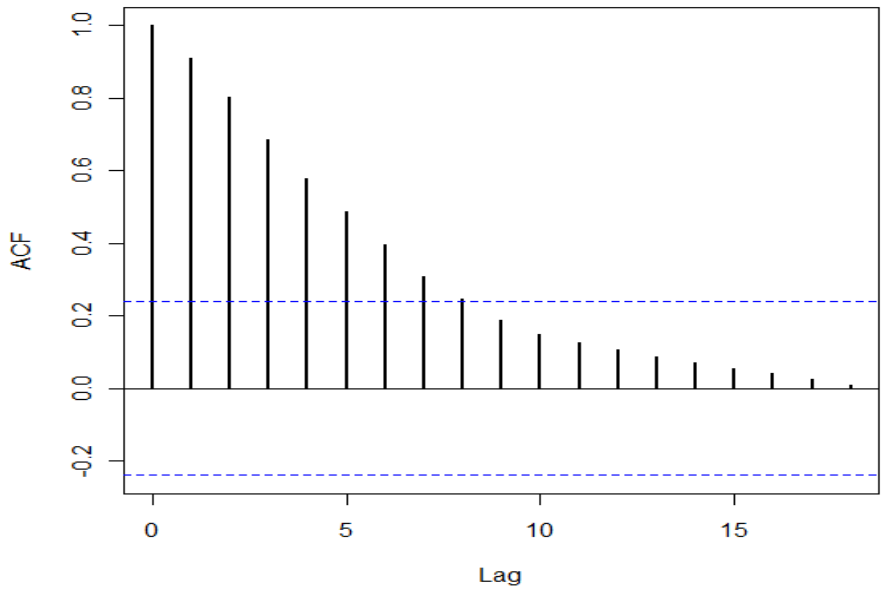


Figure-2 Plot of ACF of Imports of Pakistan



3.3 Estimation and Diagnosis for Imports

We entertained ten tentative ARIMA models as shown in the table-2, with R a function *arima (data, order (p, d, q))* from the ts–library is used to estimate the parameters and choose the model which has minimum AIC. The models and correspond log-likelihood along with the AIC can be seen in table-2 given below.

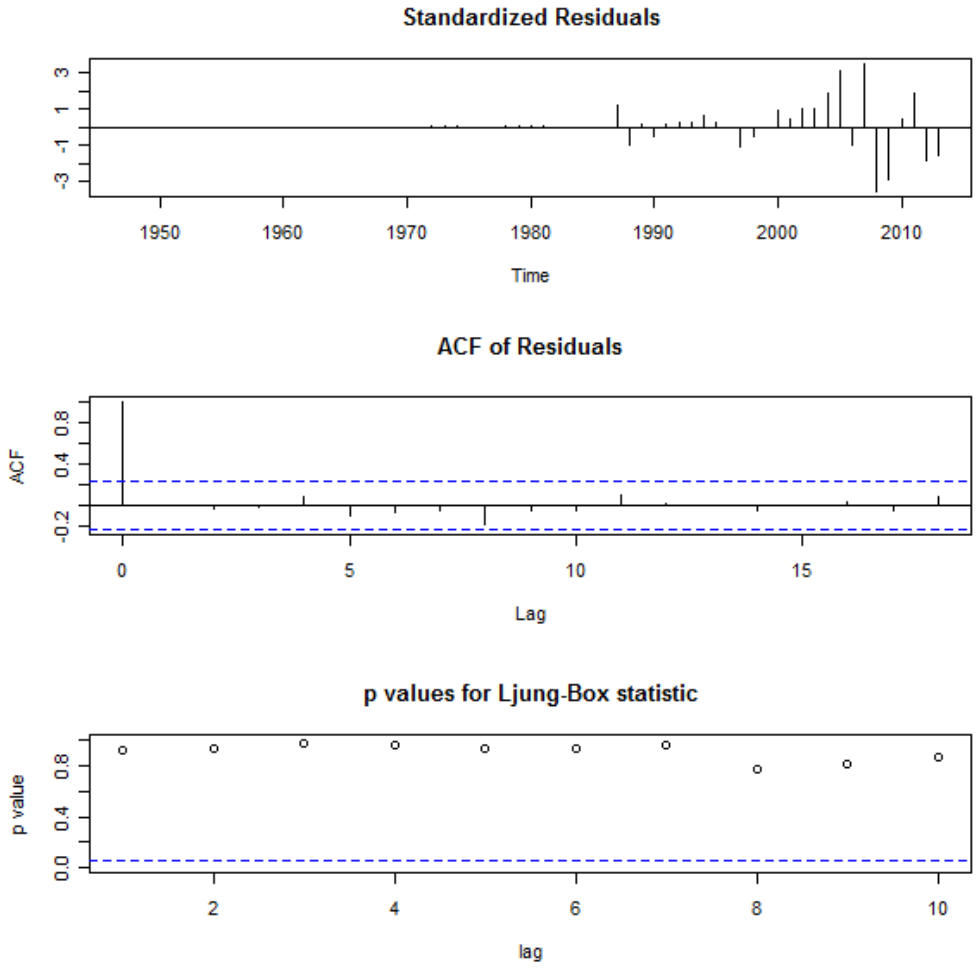
Table 2: Tentative Models for Imports of Pakistan

Models	**Log likelihood**	**AIC**
ARIMA(0, 1, 0)	-730.36	1462.72
ARIMA(1, 1, 0)	-715.39	1434.78
ARIMA(1, 1, 1)	-711.11	1428.55
ARIMA(0, 1, 1)	-720.71	1445.43
ARIMA(1, 2, 1)	-701.40	1408.79
ARIMA(2, 2, 1)	-698.71	1405.42
ARIMA(2, 2, 2)	**-692.24**	**1394.48**
ARIMA(1, 2, 2)	-700.99	1409.97
ARIMA(2, 2, 1)	-698.71	1405.42
ARIMA(0, 2, 0)	-711.77	1425.54

R has the function *tsdiag(fit),* which produces a diagnostic plot of a fitted time series model after running the code *fit<-arima(data, order=c(p, d, q)),*and *tsdiag(fit).*The table-4 and the plots given below tell us the p-values for the Ljung-Box-Pierce statistics for each lag up to 10. These statistics consider the accumulated residual autocorrelation from lag 1 up to and including the lag on the horizontal axis. The dashed blue line is at .05. All p-values are above it. That's a good result. We want non-significant values for this statistic when looking at residuals.

Table 3: Box Pierce test for Residuals of Fitted ARIMA (2, 2, 2) Model

Chi square	Df	p-value
4.6404	10	0.9139



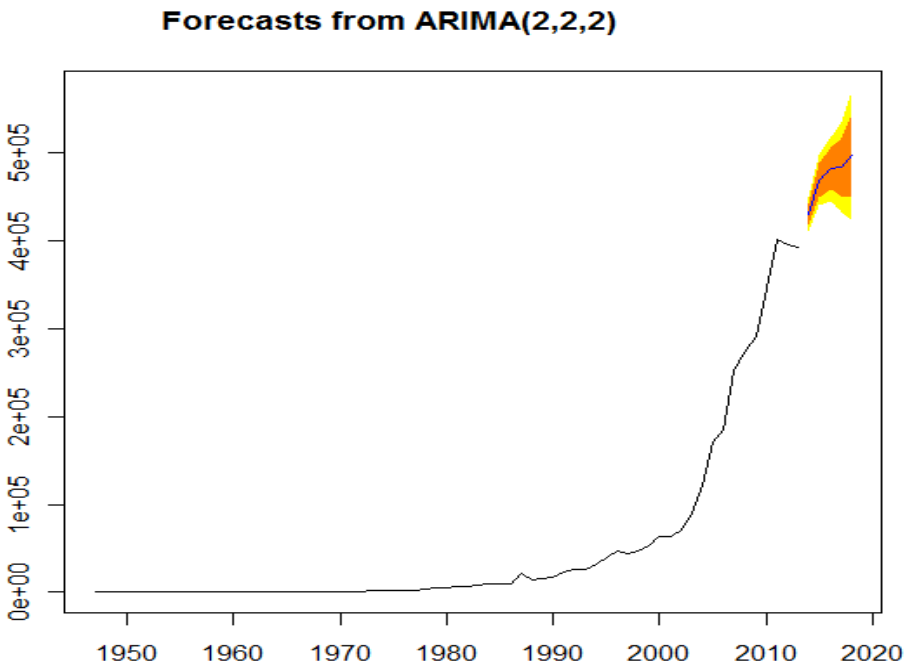
3.4 Forecasting for Imports

Once a model has been identified, its parameters estimated and diagnostic have been checked, one purpose is to forecast the future values of a time series. We can then use the fitted ARIMA model to make forecasts for future values of the time series, using the *forecast.arima ( )* function in the forecast package.

Table 4: Forecasted values in million PKR

Years	Forecast
2014	429955.1
2015	468763.5
2016	481785.7
2017	483424.8
2018	496956.5

The plot given below gives us an observed Imports values for the years 1947 to 3013, as well as the Imports that would be predicted for these 67 years and for the next 5 years forecasted values of Imports using our ARIMA (2, 2, 2) model, by typing a code *plot.forecast( )* in R.



3.5 Summary Statistics for Exports

The summary statistics for Exports are given below in the table-5 again by using R with function *describe (data*) from the psych-library.

Table 5: Summary Statistics of Exports of Pakistan

N	Mean	Median	Standard deviation	Min	Max	Skewness	Kurtosis	Standard error
67	32976.9	2341.01	61633.03	43.36	275917	2.25	4.42	7529.67

It can be seen from the graph in figure-3 that, the Exports of Pakistan is also increasing and decreasing slowly over time up till year 2008 and after that it was continuously increasing up to year 2013. The plot of autocorrelation in figure-4 shows that the sample autocorrelations are very strong, positive and decay slowly as we already seen in Imports of Pakistan which also indicates that there are possible shifts in both the mean and the variability over time for this series and the trend can be removed by differencing once or twice. Again an untransformed difference is used in future analyses.

Figure-3 Graph of Exports of Pakistan

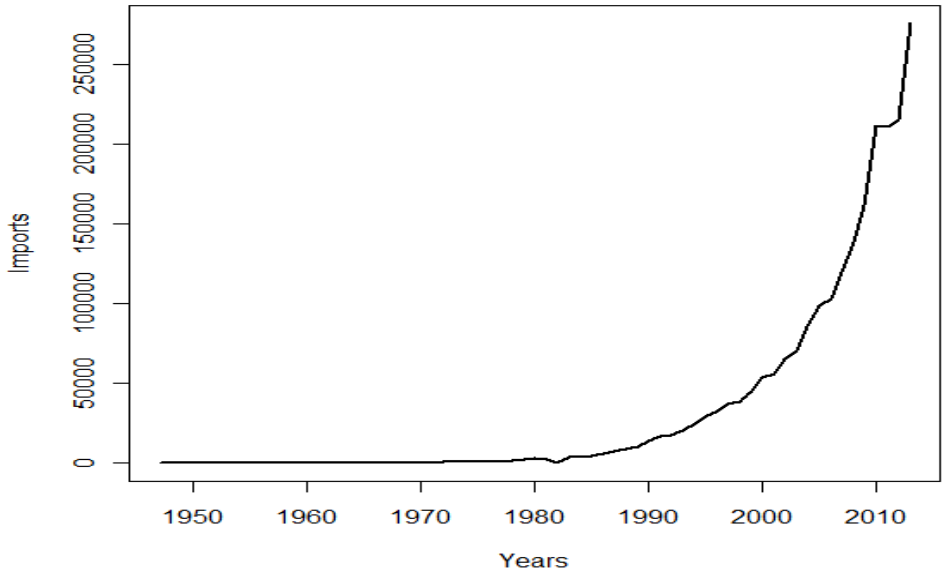
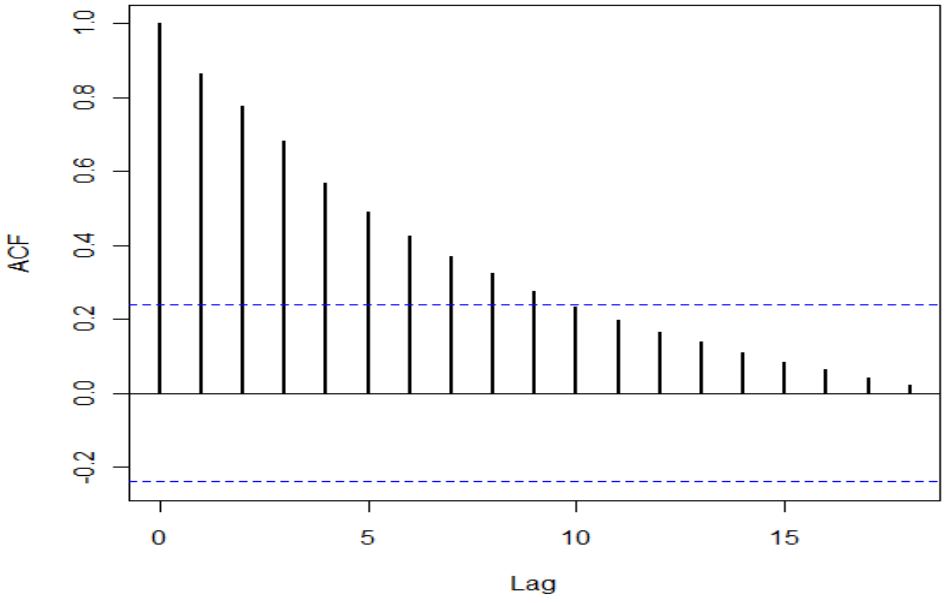


Figure-4 Plot of ACF of Exports of Pakistan



3.6 Estimation and Diagnosis for Exports

Again we entertained 10 tentative ARIMA models as shown in the table-6, by using function *arima (data, order (p, d, q))* from the ts–library in R to estimate the parameters and choose the model which has minimum AIC again. The models and correspond log-likelihood along with the AIC can be seen in table-6 given below.

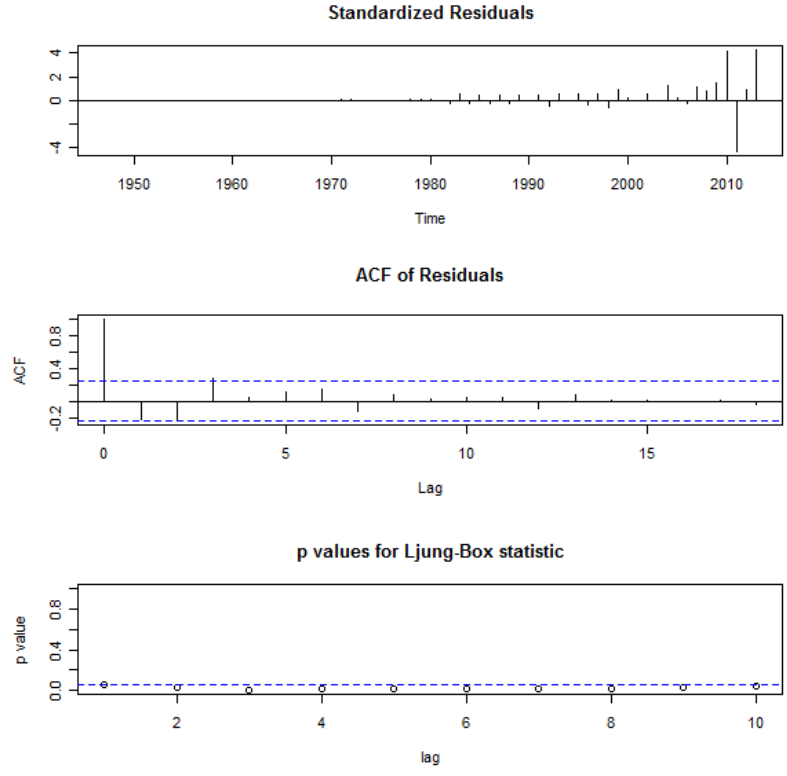
Table 6: Tentative Models for Exports of Pakistan

Models	Log likelihood	AIC
ARIMA(0, 1, 0)	-708.64	1419.28
ARIMA(1, 1, 0)	-701.17	1406.33
ARIMA(1, 1, 1)	-696.18	1398.36
ARIMA(0, 1, 1)	-696.95	1397.89
ARIMA(1, 2, 1)	-679.88	1365.76
ARIMA(0, 2, 0)	-693.99	1389.99
ARIMA(1, 2, 2)	**-674.55**	**1357.11**
ARIMA(0, 2, 0)	-693.99	1389.99
ARIMA(0, 2, 1)	-680.94	1365.89
ARIMA(0, 2, 2)	-681.11	1368.21

Using same function *tsdiag(fit),* in R produces a diagnostic plot of a fitted time series model after running the code *fit<-arima(data,order=c(p,d,q)),* and *tsdiag(fit).*The table-7 and the plots given below tell us the p-values for the Ljung-Box-Pierce statistics for each lag up to 10 All p-values are marginally on it which is not a very good result. We have marginally non-significant values for this statistic when looking at residuals of the fitted model for Exports.

Table 7: Box Pierce test for Residuals of Fitted ARIMA (1, 2, 2) Model

Chi square	Df	p-value
17.1785	10	0.07015



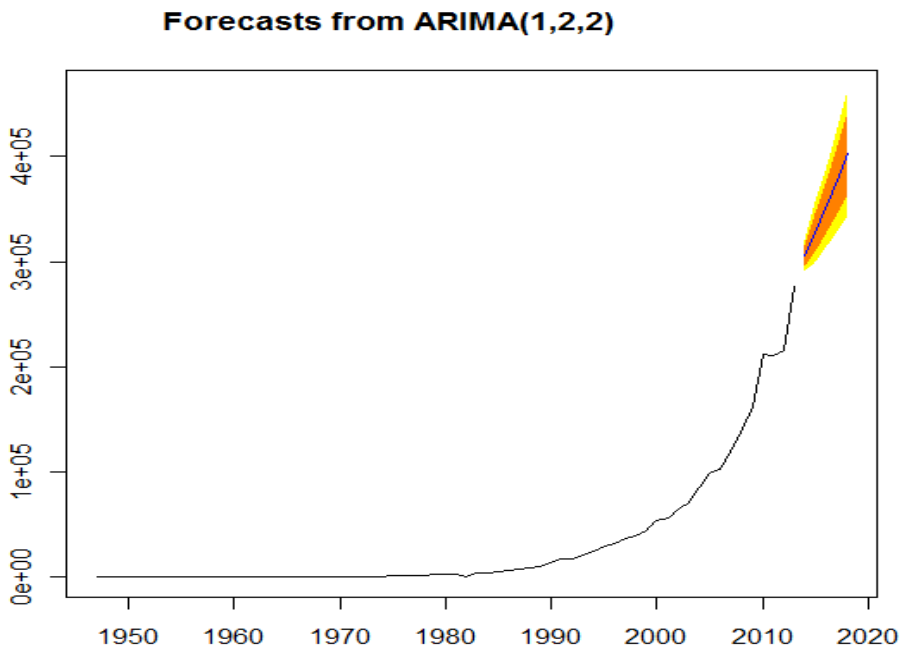
3.7 Forecasting For Exports

Again after a model has been identified, its parameters estimated and diagnostic have been checked, next step is to forecast the future values of a time series. Again we have used the fitted ARIMA model to make forecasts for future values of the time series, using the *forecast.arima ( )* function in the forecast package.

Table 8: Forecasted values in million PKR

Years	Forecast
2014	305296.4
2015	327518.9
2016	353830.9
2017	377806.1
2018	403116.5

The plot given below gives us an observed Exports values from 1947 to 2013, as well as the Exports that would be predicted for these 67 years and for the next 5 years forecasted values of Exports using our ARIMA (1, 2, 2) model, by typing a code *plot.forecast()* in R.



4. General Conclusion and Suggestions

An ARIMA model offers a good technique for forecasting the magnitude of any time dependent variable. Its strength lies under the fact that the method is suitable for any time series data with any pattern of change and is suitable for at least 50 observations which is one of its limitations. The basic aim of the study was to select models for forecasting Import and Export of Pakistan. In this context, we took our interest on ARIMA with respect to our data. It was found that AIC based model selection procedures gave ARIMA models with order (2, 2, 2) and order (1, 2, 2) for Imports and Exports of Pakistan respectively were appropriate. It should also be borne in mind that a good forecasting technique for a situation may become inappropriate technique for a different situation. The validation of particular model must be examined as time changes. The researcher can use these models for forecasting annual Imports and Exports of Pakistan. However it should be updated time to time with an incorporation of current data. It is most important that the researcher should always follow the principle of parsimony and try to fit simple model instead of model with large parameters.

References

1. Box, G.E.P and Jenkin, G. M (1976). *Time series analysis in forecasting and control Applied statistics*. Holden-Day, San Francisco.
2. Challis, R. E., and Kitney, R. I. (1991). Biomedical signal processing, Part 1 Time-domain methods. *Medical and Biological Engineering and Computing*, 28, 509-524.
3. Economy Watch. March 29, 2010, http://www.economywatch.com/world_economy/pakistan/export- mport.html (accessed March 31, 2013).
4. Granger, C. W. J. and Newbold, P. (1976). Forecasting transformed series, *Journal of the Royal Statistical Society B*, 38, 189-203.
5. Kendall, M. G (1995). *Time series.* Griffin and Co ltd, London.
6. Ljung, G.M. and Box, G.E.P. (1978). On a Measure of Lack of Fit in Time Series Models, *Biometrika*, 65, 297-303.
7. Shair, Ali (2006). *Effects of major exports and imports on the balance of foreign trade in Pakistan, p.4,* Master's Thesis, Lund University.
8. Zaidi, S. Akbar, (2000). *Issues in Pakistan Economy,* Oxford University press.
9. 50 years of Pakistan (volume) (1947-1997). *Federal Bureau of Statistics, Government of Pakistan*.