India’s GDP-The Next Decade: A Forecast Using ARIMA Model

Dr. Ravi Changle*
Dr. Sukhjeet Matharu**
Arnav Choudhary* **
Gaurav Malhotra****

***Assistant Professor, Prestige Institute of Management and Research, Indore**
****Assistant Professor, Prestige Institute of Management and Research, Indore**
*****Assistant Professor, Prestige Institute of Management and Research, Indore**
******Assistant Professor, Indore Indira Institute of Management, Indore**

Abstract

India’s growth along with sustainability in the next decade majorly depends on the growth in its market and economy as a whole. In the present study researcher has attempted to forecast the GDP growth for the country. Out of a variety of forecasting models ARIMA (1,2,2) model has been applied to forecast the GDP over a period of ten years ranging from 2015 to 2025. The results indicate the fitness of AR (1) I (2) MA(2) parameters for making the future predictions. It is concluded that the GDP of India would be rising continuously over the estimated period. Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) were studied to study the AR and MA terms. Augmented Dickey Fuller (ADF) Unit root test was used to test the stationarity of the data and identification of Integration (I) order. Akaike Information Criterion (AIC), Root Mean Squared Error, Mean Absolute Percentage Error were applied to study the model fitness.

Keywords: *Auto Regressive Integrated Moving Average (ARIMA), Autocorrelation Function (ACF), Akaike Information Criterion (AIC), Forecast, GDP, Mean Absolute Percentage Error, Partial Autocorrelation Function (PACF), Root Mean Squared Error*

Introduction:

The assessment of the current state of the economy is an important element in macroeconomic forecasting for a longer term analysis. Researchers in economics are particularly interested in GDP forecast for assessing and predicting the functional status of the economy of the developing countries all over the world. GDP indicates the financial health of a country as a whole-which is actually a hunting ground of researchers in the field of business in general and of economics in particular. The issues of GDP has become the most concerned amongst macro economy variables and data on GDP is regarded as the important index for assessing the national economic development and for judging the operating status of macro economy as a whole (Ning et al., 2010).

The GDP of a country serves as a basis for setting and planning of the economic developmental policies. Researchers and policy makers are interested in gaining an insight into the future health of an economy. This is made possible with the help of suitable sophisticated time series modelling techniques. These techniques provide an accurate prediction of GDP which is helpful to the policy makers in general. Industries in general and individual firms are also interested in the forecast of GDP as it serves as a basis for planning the investment and expansion strategies. It also provides a basis for framing strategic development policies, economic policies and deciding the allocation of funds in the priority sector.

GDP is the aggregate statistic of all economic activity and captures the broadest coverage of the economy than other macro economic variables. It is the market value of all final goods and services produced within the borders of a nation in a year. It is often considered the best measure of how well the economy is performing. GDP can be measured in three ways. First, the Expenditure approach, it consists of household, business and government purchases of goods and services and net exports. Second the Production approach, it is equal to the sum of the value added at every stage of production (the intermediate stages) by all industries within the country, plus taxes and fewer subsidies on products in the period. Third is Income approach, it is equal to the sum of all factor income generated by production in the country (the sum of remuneration of employees, capital income, and gross operating surplus of enterprises i.e. profit, taxes on production and imports less subsidies) in a period (Yang, 2009- 2010).

In a study, Tsay and Tiao (1984, 1985) used ARIMA model, which is in fact fitted on non-seasonal data by identifying autoregressive and moving average terms with the help of partial autocorrelation and autocorrelation functions (Box and Jenkins, 1970:1976,Pankratz, 1991).However, in the case of seasonal data, a number of studies used filtering approach, which in fact very helpful in case of weekly, monthly, quarterly and semiannual data to estimate a model to forecast any macro variable (Liu, 1989; Liu and Hudak , 1992).

In another research, Reynolds et al., (1995) developed automatic methods to identify as well as estimate the parameters of ARIMA model by utilizing time-series data for a single variable. In another study, Reilly (1980) used similar methodology to model macroeconomic variable like GDP. However, both the studies confined themselves only on non-seasonal time series data and restrained to predict the variable in future. However, the above mentioned methods need a long time-series data on the macroeconomic variable in question. To estimate the model for prediction of a macro variable, a number of studies imply analytical neural network techniques, which is very effective in the case of seasonal data (Chiu et al., 1995; Cook and Chiu, 1997; Geo et al., 1997; Saad et al., 1998). These types of models have got pace since the seminal paper of (Granger and Joyeux, 1980 and Hosking, 1981. However, this neural networking approach is very difficult to applying in real life situation by the policy makers /managers due to difficult network design, training and testing are required to build the model as well as to estimate the parameters.

The researcher was motivated to undertake this study dealing with the GDP issues in India by various studies reported in the western world. Further, not many studies have been done attempting to forecast the GDP as well as prediction of growth rates in various forms in India. However an attempt to forecast this macro variable only as point estimates has been of very little help for the managers and policy makers since variability is the key in decision making when certain level of risk is involved.

In the above backdrop, this study is a modest attempt to fill the gap by identifying the following two research questions with respect to GDP forecasting issues in India

1. What are the year-wise forecasted GDP values in various forms over a period 2013-2025?
2. What are the year-wise GDP growth rates in different forms over a period 2013-2025?

Literature Review:

Bipasha Maity et al. (2012) conducted the same study for a period till 2020 using ARIMA Model. Reynolds et al. (1995) developed automatic methods to identify as well as estimate the parameters of ARIMA model by utilizing time-series data for a single variable.

Reilly (1980) used similar methodology to model macroeconomic variable like GDP. However, the studies confined themselves only on non-seasonal time series data and restrained to predict the variable in future. Tsay and Tiao (1984, 1985) used ARIMA model, which is in fact fitted on non-seasonal data by identifying autoregressive and moving average terms with the help of partial autocorrelation and autocorrelation functions (Box and Jenkins 1970:1976,Pankratz 1991).

However, the above mentioned methods need a long time-series data on the macroeconomic variable in question. To estimate the model for prediction of a macro variable, a number of studies imply analytical neural network techniques, which is very effective in the case of seasonal data (Chiu et al. 1995; Cook and Chiu 1997; Geo et al. 1997; Saad et al. 1998).

These types of models have got pace since the seminal paper of Granger and Joyeux (1980) and Hosking (1981). However, this neural networking approach is very difficult to applying in real life situation by the policy makers /managers due to difficult network design, training and testing are required to build the model as well as to estimate the parameters.

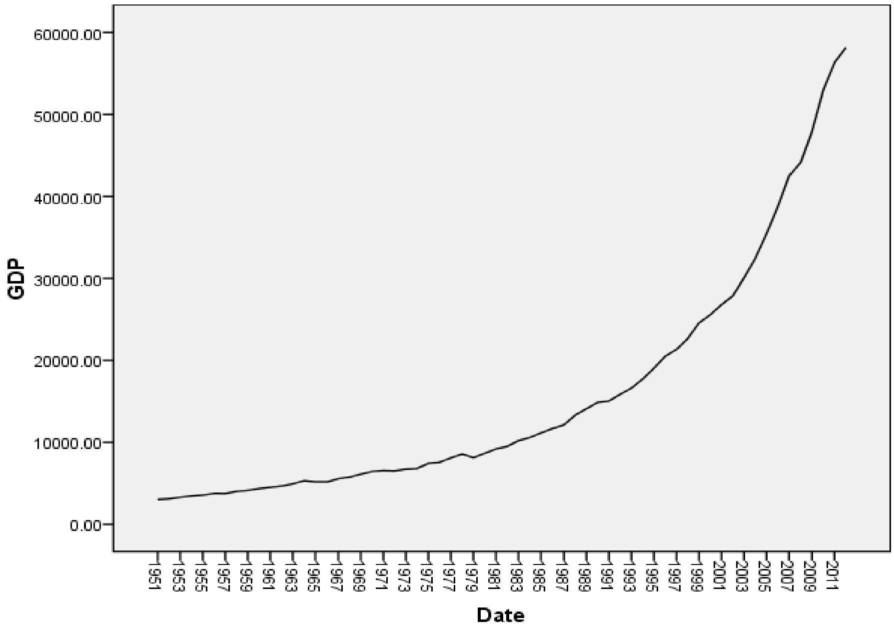
Research Objectives:

- To test the stationarity in the data of GDP over the period.
- To study Autocorrelation in the observed series of GDP
- To Forecast the GDP using appropriate ARIMA Model.
- To test the Model fitness using Information Criterion and Badness of fit model.

The Data:

The data of GDP (Nominal Value) was collected over the time period from 1951 to 2013. It had been collected from the website publications of Reserve Bank of India (RBI). The following figure exhibits the GDP Curve over the observed period of 62 years.

Figure-1: GDP of India in Rs. billion



Conceptual Framework of ARIMA Model:

The publication of *Time Series Analysis: Forecasting and Control* by Box and Jenkins ushered in a new generation of forecasting tools. Popularly known as the Box-Jenkins (BJ) methodology, and technically known as the ARIMA methodology, the emphasis of these methods is not on constructing single equation or simultaneous equation models but on analyzing the probabilistic or stochastic properties of economic time series on their own under the philosophy of letting the data speak for themselves. Unlike the regression models, in which *Yt* is explained by *k* regressors, *X*1, *X*2, *X*3...*Xk*, the BJ type time series models allow *Yt* to be explained by past or lagged valued of *Yt* itself and stochastic error terms. For this reason, ARIMA models are sometimes called a-theoretic models because these are not derived from any economic theory, while economic theories are often the basis of simultaneous equation models.

Let Y_t be a time series sequence for t = 1, 2 ….t as:

$$y_t - \delta = \alpha_1 (y_{t-1} - \delta) + u_t$$

Where $\boldsymbol{\delta}$ is the mean of *Yt* and where *ut~ ii d N* (0, σ2 ε), then we can say that *Yt* follows a first order Autoregressive (AR)(1). Here the value of *Y* at time *t* depends on its value in the previous time period and a random term. In other words, this model says that the forecast value of *Y* at time *t* is simply some proportion (α_1) of its value at time (*t* – 1) plus a random shock or disturbance at time *t*, again the values are expressed around their mean values. Economic variables with time series data are usually non-stationary, since these are integrated. These need first order differencing for attaining stationarity.

If a time series is integrated of order 1, its first differences are *I* (0), and it is stationary. Similarly, if a time series is *I* (2), its second difference is *I* (0). In general, if a time series is *I* (*d*), after differencing it *d* times, we obtain an I (0) series. Therefore , if we have to difference a time series *d* times to make it stationary and then apply an ARIMA time series, where *p* denotes the number of AR terms, *d* represents the number of times, the series has to be differenced before it becomes stationary, and *q* is the number of MA terms.

Results:

Identification of I (d) term:

The GDP series was differences twice as ADF test (Table-1) justified the presence of unit root in the data. Thus, the stationarity was achieved by differencing the GDP series twice. It thus, validated the integration order at two lags I(2).

Table-1: ADF Unit Root Test

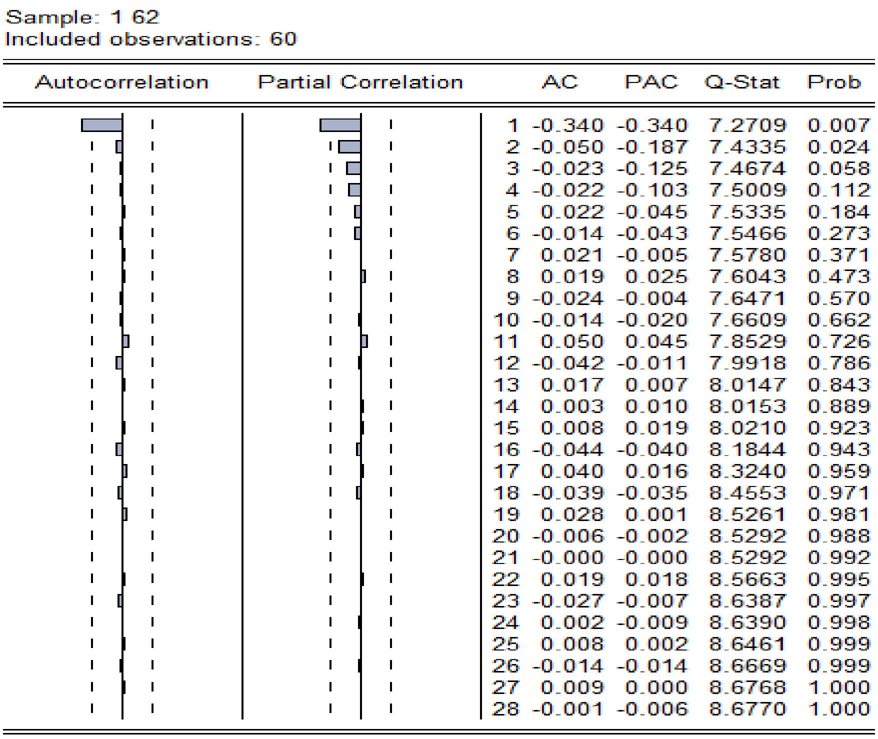
		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		4.417378	1.0000
Test critical values:	1% level	-3.568308	
	5% level	-2.921175	
	10% level	-2.598551	

*MacKinnon (1996) one-sided p-values.

Identification of AR (p) and MA (q) terms:

After making the GDP series stationary the autocorrelation and partial autocorrelation functions were studied. By observing the PACF (figure-2) values and correlogram term AR (1) was found to be fit for predictions. Similarly, MA (1) and (2) terms were justified by observing the ACF (figure-2) values and correlogram. The MA (1) was rejected as it was found to be insignificant. Thus, the ARMA (1, 2) parameters were identified using Autocorrelation and Partial Autocorrelation Functions.

Figure-2: Correlogram for ACF and PACF Values



The ARIMA (1, 2, 2) parameters (Table-2) were found significant at 5% level of significance. The coefficient of AR (1) was estimated as 0.54 and that of MA (2) as 0.49. The t-test confirms the significance of these coefficients for predicting the GDP. A model ARIMA (1, 1, 1) was rejected due to insignificance of its AR (1) term while testing. The model fitness was confirmed by lower AIC values and lower values of root mean squared error. The R-square is merely 24%.

Table-2: ARIMA (1, 2, 2) Model

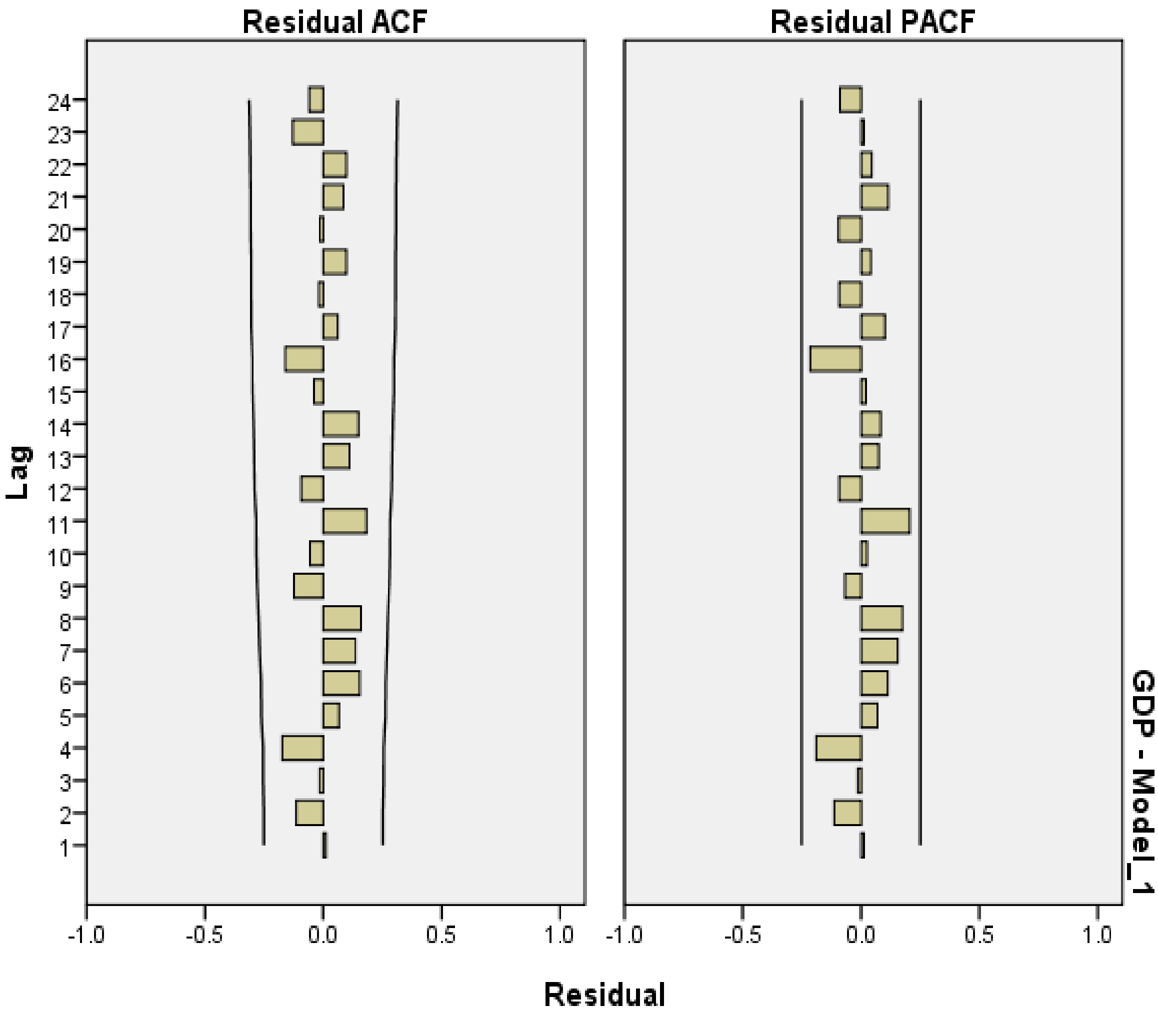
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	51.31709	26.19999	1.958669	0.0551
AR(1)	-0.541294	0.134186	-4.033904	0.0002
MA(2)	-0.491852	0.138117	-3.561114	0.0008

R-squared	0.240660	Mean dependent var	27.65864
Adjusted R-squared	0.213541	S.D. dependent var	661.5928
S.E. of regression	586.7171	Akaike info criterion	15.63647
Sum squared resid	19277270	Schwarz criterion	15.74211
Log likelihood	458.2759	F-statistic	8.874147
Durbin-Watson stat	1.875823	Prob(F-statistic)	0.000449

Model Validity:

The correlogram of ACF of residuals (Fig 3) suggests that there is no substantial spike has been observed as the case of correlogram of PACF of residuals. In turn, it has been concluded that the error terms become white noise. It thus validates the model as no further information is available.

Figure-3: Residual ACF and PACF



To test the model validity statiscally a portmanteau test of Independence was applied. The BDS test (Figure-3) is a portmanteau test for time based dependence in a series. It was used for testing against a variety of possible deviations from independence including linear dependence, non-linear dependence, or chaos.

The test can be applied to a series of estimated residuals to check whether the residuals are independent and identically distributed (iid). For example, the residuals from an ARMA model can be tested to see if there is any non-linear dependence in the series after the linear ARMA model has been fitted. The idea behind the test is fairly simple. To perform the test, we first choose a distance. We then consider a pair of points. If the observations of the series truly are iid, then for any pair of points, the probability of the distance between these points being less than or equal to epsilon will be constant. The BDS statistic too confirms the statistical validity of the model.

Figure-3: BDS Independence Test

Dimension	BDS Statistic	Std. Error	z-Statistic	Normal Prob.	Bootstrap Prob.
2	0.057478	0.014857	3.868692	0.0001	0.0040
3	0.067920	0.023984	2.831878	0.0046	0.0204
4	0.067551	0.029028	2.327068	0.0200	0.0466
5	0.034732	0.030763	1.129005	0.2589	0.2384
6	-0.006853	0.030174	-0.227109	0.8203	0.9326

The GDP Forecast:

Table-3: GDP Forecast using ARIMA (1,2,2)

GDP In Rs. billion

Year	GDP_Forecast	Growth rate	UCL	% change	LCL	% change
2013	64796.42	-	65658.97	-	63933.88	-
2014	69174.76	6.75706	70860.89	7.92264	67488.62	5.56003
2015	73587.93	6.37975	76116.91	7.41738	71058.96	5.29028
2016	78115.88	6.15311	81636.34	7.25126	74595.42	4.9768
2017	82690.56	5.85627	87254.42	6.88183	78126.69	4.7339
2018	87369.89	5.65885	93092.65	6.69103	81647.13	4.50606
2019	92104.56	5.41911	99042.2	6.391	85166.92	4.31098
2020	96936.55	5.24621	105183.4	6.2006	88689.7	4.13632
2021	101830.1	5.04822	111442.3	5.95046	92217.95	3.9782
2022	106815.7	4.89598	117873.6	5.77092	95757.87	3.83864
2023	111867.4	4.72935	124425.9	5.55877	99308.93	3.70837
2024	117007.3	4.59459	131136.9	5.39361	102877.6	3.59348
2025	122216.5	4.45206	137970.7	5.21116	106462.3	3.48444

The nominal values of GDP along with Upper and lower control levels were forecasted using ARIMA (1, 2, 2) model. Though the nominal values are increasing exponentially, the growth rates were found to be declining over the estimated period. Higher growth rates were expected in the period of 2014 to 2016, thereafter the growth rates were found to be stagnant and GDP is supposed to increase with a declining rate of growth. It is thus suggested that for making FDI or FII decision it should be imperative to foresee growth rates apart from nominal GDP values.

Conclusion:

The GDP forecast for the next decade though is exhibiting an inclining trend nominally but the growth rate is declining over the period of time. It thus signifies the role of RBI to take appropriate measures regarding the monetary policy and credit policy so that the growth is not hampered in the future. Recent reports from United Nations is overlooking China's growth rate and predicts the Indian economy would be overhauling the Chinese growth in the Asian markets, but the results as per a-theoritic model developed by Box-Jenkins is revealing the meagre growth rates for the next decade. The ARIMA (1, 2, 2) model was found to be a better fit model in forecasting India's GDP.

Implications of the Study:

The results of this study would be very useful for policy makers and managers dealing with macro variables such as foreign direct investment (FDI), foreign institutional investment (FII), etc. The findings of the study will be helpful for formulation of better policies. Managers who are planning to invest in the expansion of existing business or in the new project will be benefitted greatly as the findings will provide them a picture of the economic conditions of India well in advance. Further, the findings suffer from some limitations since the researchers have not taken into consideration the models such as Regression Analysis, VAR, ECM etc. to forecast GDP and its growth rates in India

References:

- Ansley, C.F. (1979), An algorithm for the exact Likelihood of a mixed autoregressive-moving average process, *Biometrika*, 66: 59-65.
- Ard, H.J, Den, R. (2010), Macroeconomic Forecasting using Business Cycle leading indicators, Stockholm: US-AB.
- Bipasha, M. and Chatterjee, B. (2012), Forecasting GDP Growth Rates of India: An Empirical Study. *International Journal of Economics and Management Sciences*, 1(9): 52-58.
- Box, G.E.P. and Jenkins, G.M. (1970), Time Series Analysis: Forecasting and Control, San Francisco: Holden-Day.
- Box, G.P. and Jenkins, G.M. (1976), Time Series Forecasting and Control, San Francisco: Holden-Day.
- Box, G.E.P., Jenkins, G.M. and Reinsel, G.C. (1994), Time Series Analysis: Forecasting and Control, Englewood Cliffs, NJ: Prentice Hall.
- Chiu, C.C., Cook, D.F. and Pignatiello, J.J. (1995), Radial basis function neural network for kraft pulping forecasting, *International Journal of Industrial Engineering*, 2(2). 209-215.

- Cook D.F. and Chiu, C.C. (1997), Predicting the internal bond strength of particleboard utilizing a radial basis function neural network, Engineering Applications AI, 10(2):171-177.
- Gao, X.M., Gao, X.Z., Tanskanen, J. and Ovaska, S.J. (1997), Power prediction in mobile communications systems using an optimal neural structure, IEEE Transportation Neural Networks, Vol. 8 No. 6, pp. 1446-1455.
- Granger, C.W.J. and Joyeux R. (1980), An introduction to long-memory time series models and fractional differencing, *Journal of Time Series Analysis*, 1:15-39.
- Enders, W. (2004), Applied Econometric Time Series, New York: John Wiley and Sons.
- Hoff, J.C. (1983), A Practical guide to Box-Jenkins Forecasting, London: Lifetime Learning Publications.
- Hosking J.R.M. (1981), Fractional differencing, Biometrica, 68: 165-176.
- Hossain, M.Z., Samad, Q.A. and Ali, M.Z. (2006), ARIMA model and forecasting with three types of pulse prices in Bangladesh: A case study, International Journal of Social Economics, 33(4): 344-353.
- Liu, L.M. (1989), Identification of seasonal ARIMA models with using a filtering method, Communications in Statistics, 18: 2279-2288.
- Liu, L.M. (1999), Forecasting and time series analysis using the SCA Statistical System: Vol. 2, Chicago: Scientific Computing Associates Corp.
- Liu, L.M. and Hudak, G.B. (1992), Forecasting and time series analysis uses the SCA Statistical System: Vol. 1 Chicago: Scientific Computing Associates Corp.
- Ljung, G.M. and Box, G.E.P. (1978), On a measure of lack of fit in time series models, Biometrika, 65(1): 297-303.
- McDowall, D., McCleary, R., Meidinger, E.E. and Hay, R. A. (1980), Interrupted time series Analysis, Baverly Hills, CA: Sage Publications.
- Melard, G. (1984), A fast algorithm for the exact likelihood of auto regressive- moving average models, Applied Statistics, 33:104-119.
- Ning, W., Kuan-jiang, B. and Zhi-fa, Y. (2010), Analysis and forecast of Shaanxi GDP based on the ARIMA Model, *Asian Agricultural Research*, 2(1): 34-41.
- O'Donovan, T.M. (1983), Short Term Forecasting: An Introduction to the Box-Jenkins Approach, New York: John Wiley and Sons.
- Pankratz, A. (1983), Forecasting with Univariate Box-Jenkins models. Concepts and Cases, New York: John Wiley and Sons.

- Pankratz, A. (1991), Forecasting with dynamic regression models, New York: John Wiley and Sons.
- Reilly, D.P. (1980), Experiences with an automatic Box-Jenkins modeling algorithm. Time Series Analysis- Proceedings of Houstan Meeing on Time Series Analyis, Amsterdam: North-Holland publishing.
- Reynolds, S.B., Mellichamp, J.M. and Smith, R.E. (1995), Box-Jenkins forecast model identification. AI Expert, June, pp. 15-28.
- Saad, E.W., Prokhorvo, D.V. and Wunsch, D.C. (1998), Comparative study of stock trend prediction using time delay. Recurrent and probabilistic neural networks, IEEE Transportation Neural Networks, 9(6):1456-1469.
- Tsay, R.S. and Tiao, G.C. (1984), Consistent estimates of autoregressive parameters and extended sample autocorrelation function for stationary and non-stationary ARMA models, Journal of American Statistical Association, 79: 84-96.
- Tsay, R.S. and Tiao, G.C. (1985), Use of canonical analysis in time series model identification, Biometrika, 72: 299-315.
- Yang, Lu. (2009), Modeling and a forecasting China’s GDP data with time series model. Thesis unpublished.
- Vandaele, W. (1983), Applied time series and Box-Jenkins models, New York: Academic Press.