A Study Forecasting the GDP of India Using ARIMA Model

Ravi Changle*

Dr. Sukhjeet Matharu*

Abstract

In this study one of the most prominent macroeconomic indicators is forecasted for a period till 2020 which is none other than Gross Domestic Product. The historical data for India's GDP is collected from the year 1951 and the growth rates are predicted for the same period on yearly basis. An attempt has been made to apply a-theoretic model for forecasting the Indian GDP and its growth rate i.e., ARIMA (Autoregressive Integrated Moving Average Model) and to evaluate the model's accuracy for the same. It was found that the Indian GDP's potential to grow is higher than what is observed due to adverse reactions. The research will be helpful for identifying the India's potential to grow at macroeconomic front.

Keywords: *Auto correlations, Stationarity, ARIMA, GDP, Econometric Models*

Introduction

Gross Domestic Product (GDP) of a country is the money value of all final goods and services produced by all the enterprises within the borders of a country in a year. It represents the aggregate statistic of all economic activity. The performance of economy can be measured with the help of GDP. There are three ways in which the GDP of a country can be measured. Firstly the Expenditure method (which is also known as consumption and investment method or Income disposable method, national income is estimated by aggregating of household, business and government purchases of goods and services and net exports. Secondly the Income method, it is the sum total of factor incomes earned by the normal residents of country during a year by working both within and outside the country. In other words it includes the compensation of employees, capital income, and gross operating surplus of enterprises i.e. profit, taxes on production and imports less subsidies. Thirdly the Value Added method, 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.

*Assistant Professor, Prestige Institute of Management & Research, Indore. E-mail- ravi_changle@pimrindore.ac.in

*Assistant Professor, Prestige Institute of Management & Research, Indore. E-mail- sukhjeet_matharu@pimrindore.ac.in

An accurate prediction of GDP is important to get an insight in to the future health of an economy. The government of a particular country can set up strategies for economic development on the basis of the GDP prediction. Accurate forecast of GDP with the help of suitable sophisticated time series modeling can provide a reliable estimate to the government for framing suitable economic development policies and taking decision for allocation of funds for government as well as individual firms in a particular industry on the basis of different priorities. There were some studies which attempted to forecast GDP only as point estimates which has very little help for the policy makers/ managers since variability is the key in decision making when a certain level of risk is involved The present study is an attempt to fill the gap of studies attempting to forecast the GDP as well as to predict the growth rates in various forms in India.

Review of Literature

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)

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).

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.

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. Bipasha Maity et al. (2012) conducted the same study for a period till 2020 using ARIMA Model.

Research Methodology

Data Collection

In this secondary research the data has been collected for Indian Gross Domestic Product at Constant Price from year 1951 to 2011 annually. The data is collected from the database of Reserve Bank of India's Website. GDP Growth rates are also calculated annually.

Data analysis tools & techniques

Extensive algebraic analyses are carried out in time series variables such as GDP and its growth rate to establish normality, Stationarity, pattern of shocks, conditional mean variance, etc. Another branch of researchers deal with the Auto Correlation Functions (ACFs) and Partial Autocorrelation Functions (PACFs) to examine the presence of Stationarity in the time series data sets. Yet another group of studies revolve around unit root testing, cointegration testing and time varying nature of the time series data. Very few studies have used ARIMA models for forecasting even in forecasting researchers tried with lower level lags and stop with either.

Conceptual Framework of ARIMA Models

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:

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

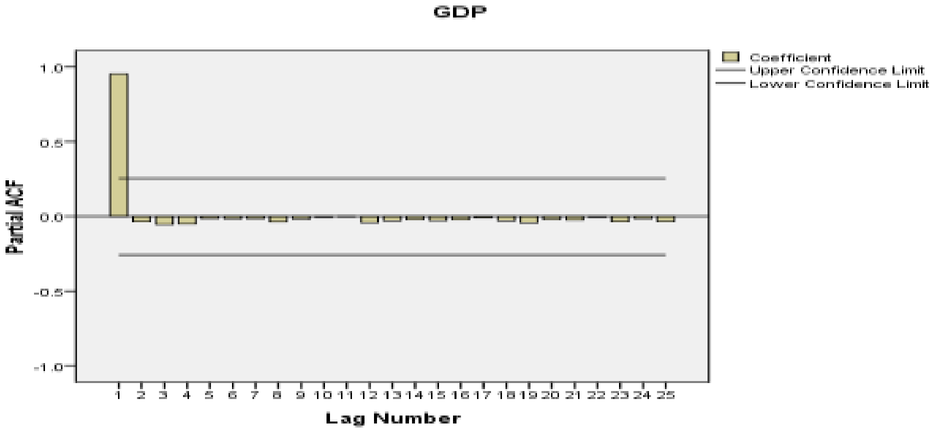
Where δ is the mean of *Yt* and where *ut~ ii d N*(0,σ^2 ε), then we can say that *Yt* follows a first order Autoregressive (AR)(1). Here the value of *Y* at time *t* depends on its value in the previous time period and a random term. In other words, this model says that the forecast value of *Y* at 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.

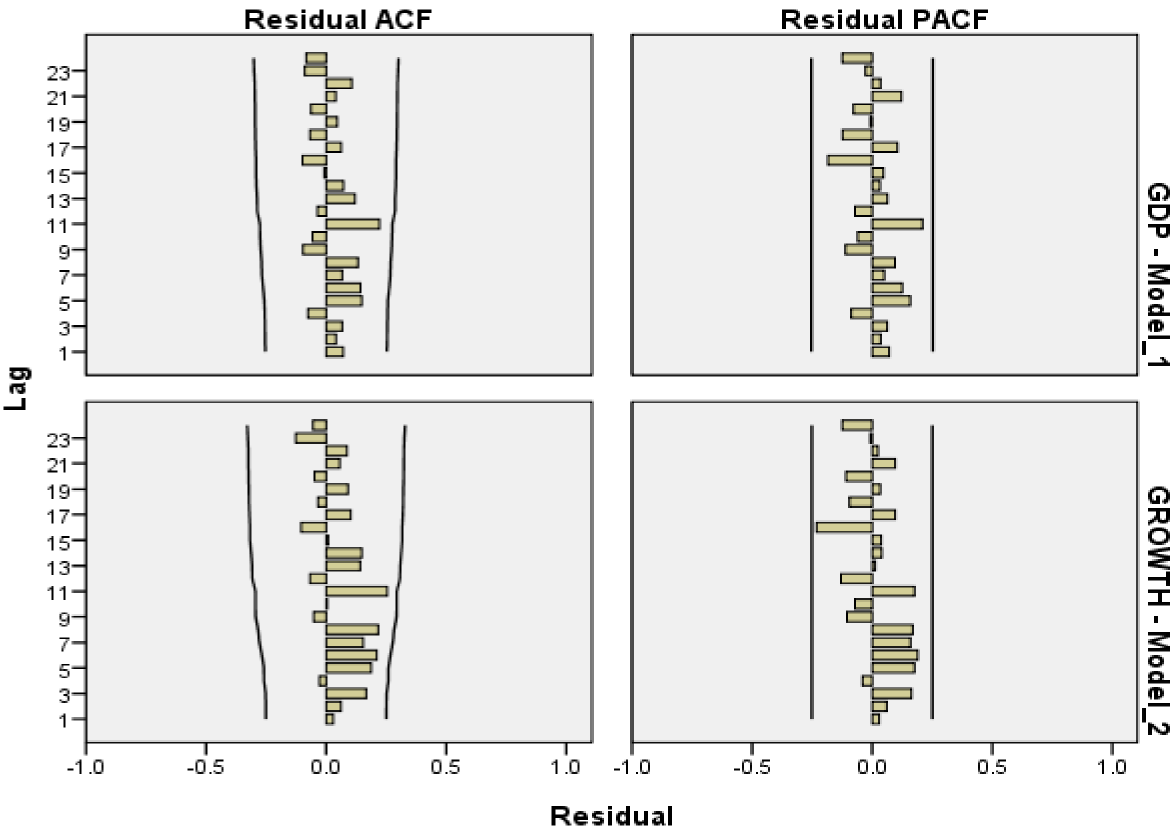
Research Objective

To forecast the Indian GDP (Constant) and its rate of growth from the year 2014 to 2020 using ARIMA Model by utilizing Time- series data over a period of 1951- 2013.

Data Analysis and Discussion

The major tools are ACF and PACF and resulting correlogram, which are simply the plots of ACFs and PACFs against the lag length[1]. In figure (1) and Table (1) & figure (2) & Table (2) the ACF and PACF Correlogram and values are shown. The ACF declines very slowly and ACF at all lags are significantly different from zero. Whereas, PACF values decline dramatically, and all the PACFs are insignificant after lag 1 (Table 2). Since the data is non-stationary it has been transformed at first difference. The results of first differenced GDP data indicates the presence of AR(1), I(1), MA(2), i.e. ARIMA (1,2,2) Model.


GDP
Coefficient
Upper Confidence Limit
Lower Confidence Limit
Partial ACF
Lag Number


Residual ACF
Residual PACF
Lag
GDP - Model_1
GROWTH - Model_2
Residual

Residual ACF

Model		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
GDP-Model_1	ACF	.049	-.159	-.221	.043	.196	-.116	-.039	-.086	-.133	-.086	.179	-.003	.006	-.093	.049	.023	-.025	-.028	-.047	-.051	.034	.117	-.025	-.020
	SE	.128	.128	.132	.137	.138	.142	.144	.144	.145	.147	.148	.151	.151	.151	.152	.152	.152	.152	.152	.153	.153	.153	.155	.155
GROWTH-Model_2	ACF	-.153	-.113	.032	-.183	.063	.111	.062	.155	-.147	-.065	.225	-.152	.090	.119	-.035	-.169	.089	-.058	.101	-.058	.074	.099	-.143	-.050
	SE	.128	.131	.133	.133	.137	.137	.139	.139	.142	.145	.145	.151	.153	.154	.155	.156	.159	.159	.160	.161	.161	.162	.163	.165

Residual PACF

Model		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
GDP-Model_1	PACF	.049	-.161	-.210	.037	.137	-.175	.036	-.059	-.227	-.126	.199	-.175	.039	.034	-.024	-.123	.070	-.128	-.088	-.038	.072	-.009	-.002	.013
	SE	.128	.128	.128	.128	.128	.128	.128	.128	.128	.128	.128	.128	.128	.128	.128	.128	.128	.128	.128	.128	.128	.128	.128	.128
GROWTH-Model_2	PACF	-.153	-.140	-.010	-.204	.000	.076	.113	.198	-.049	-.023	.231	-.062	.048	.078	.085	-.221	.069	-.101	.034	-.111	.090	.046	.024	-.086
	SE	.128	.128	.128	.128	.128	.128	.128	.128	.128	.128	.128	.128	.128	.128	.128	.128	.128	.128	.128	.128	.128	.128	.128	.128

Model Fitness

Fit Statistic	Mean	SE	Minimum	Maximum
Stationary R-squared	.483	.504	.127	.839
R-squared	.529	.662	.061	.997
RMSE	436.238	612.711	2.986	869.490
MAPE	104.000	143.272	2.691	205.309
MaxAPE	3.708E3	5.229E3	11.058	7405.896
MAE	245.435	343.730	2.381	488.489
MaxAE	2.297E3	3.235E3	9.395	4584.924
Normalized BIC	7.929	8.024	2.256	13.603

Results

The Forecast of the Indian GDP (Constant) and Growth Rate since 2014 to year 2020 using ARIMA Model is as follows:

Model		2014	2015	2016	2017	2018	2019	2020
GDP-Model_1	Forecast	67050	70316	73740	77331	81097	85046	89187
	UCL	74705	79642	84735	90025	95541	101308	107347
	LCL	60000	61833	63851	66029	68356	70826	73436
GROWTH-Model_2	Forecast	7.04	7.09	7.14	7.19	7.24	7.29	7.34
	UCL	13.07	13.15	13.22	13.30	13.38	13.46	13.54
	LCL	1.0253	1.0486	1.0709	1.0922	1.1125	1.1317	1.1499

Conclusion

The results of the study comply with the norms suggested by MLE algorithm The table-6 given below clearly indicates the presence of autocorrelations in the forex rate distribution data over the period. Further, statistical validity of the model is also checked by modified Ljung-Box statistic. In the end, results are very impressive, and the null hypothesis is rejected that there is no autocorrelation in the given data, as p-value is highly significant at 5% level of significance, and is found to be 0.000**.** As per the ARIMA Model Forecast the future of Indian GDP is optimistic. In fact this is a very economical and effective model to forecast GDP and its growth rates in India The present study will have important implications for the policy makers and the industrialists. Results of the study will be helpful for the policy makers to formulate effective policies for attracting foreign direct investment and foreign institutional investment etc. The findings of the study will also help the managerial executives to portrait a more precise picture of the economic condition of India. This will be helpful for implementing the new project ideas or taking decisions concerned with the expansion of the existing business.

Further, findings may not be best one since the researchers do not have taken into consideration of 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*, Vol. 66, pp. 59-65.

Ard, H.J, den Reijer. (2010), Macroeconomic Forecasting using Business Cycle leading indicators, Stockholm: US-AB.

Bipasha Maity and Bani Chatterjee (2012), Forecasting GDP Growth Rates of India: An Empirical Study. *International Journal of Economics and Management Sciences*, Vol. 1, No.9, 2012, pp. 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*, Vol. 2 No. 2, pp. 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*, Vol.10 No.2, pp. 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*, Vol. 1, pp. 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*, Vol. 68, pp. 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*, Vol. 33 No. 4, pp. 344-353.

Liu, L.M. (1989), Identification of seasonal ARIMA models with using a filtering method, *Communications in Statistics*, Vol. A18, pp. 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 using 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*, Vol.65 No. 1, pp. 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*, Vol. 33, pp. 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*, Vol. 2 No. 1, pp. 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*, Vol. 9 No. 6, pp. 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 ARIMA models, *Journal of American Statistical Association*, Vol. 79, pp. 84-96.

Tsay, R.S. and Tiao, G.C. (1985), Use of canonical analysis in time series model identification, *Biometrika*, Vol. 72, pp. 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.

Appendix-I

Autocorrelations

Series:GDP

Lag	Autocorrelation	Std. Error[a]	Box-Ljung Statistic		
			Value	df	Sig.[b]
1	.952	.125	58.053	1	.000
2	.903	.124	111.167	2	.000
3	.851	.123	159.190	3	.000
4	.798	.122	202.074	4	.000
5	.745	.121	240.174	5	.000
6	.694	.120	273.819	6	.000
7	.644	.119	303.348	7	.000
8	.594	.117	328.929	8	.000
9	.545	.116	350.886	9	.000
10	.499	.115	369.616	10	.000
11	.455	.114	385.521	11	.000
12	.410	.113	398.731	12	.000
13	.366	.112	409.470	13	.000
14	.323	.111	418.007	14	.000
15	.280	.109	424.570	15	.000
16	.238	.108	429.422	16	.000
17	.199	.107	432.874	17	.000
18	.160	.106	435.151	18	.000
19	.120	.105	436.461	19	.000
20	.081	.103	437.076	20	.000
21	.043	.102	437.257	21	.000
22	.009	.101	437.264	22	.000
23	-.026	.099	437.333	23	.000
24	-.059	.098	437.690	24	.000
25	-.091	.097	438.571	25	.000

a. The underlying process assumed is independence (white noise).

b. Based on the asymptotic chi-square approximation.

Partial Autocorrelations

Series:GDP

Lag	Partial Autocorrelation	Std. Error
1	.952	.128
2	-.036	.128
3	-.054	.128
4	-.050	.128
5	-.014	.128
6	-.016	.128
7	-.015	.128
8	-.037	.128
9	-.019	.128
10	-.006	.128
11	.000	.128
12	-.041	.128
13	-.030	.128
14	-.022	.128
15	-.028	.128
16	-.023	.128
17	-.010	.128
18	-.031	.128
19	-.043	.128
20	-.021	.128
21	-.025	.128
22	-.007	.128
23	-.036	.128
24	-.015	.128
25	-.035	.128