IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

A Comparison Of Time Series Models For Forecasting Gdp Of India

Prajna R

Department of statistics, Mangalore University, Karnataka

Abstract

Gross Domestic Product tracks the health of a country's economy. It is the total monetary value of all the finished goods and service product within the country in a specific period of time, usually a year or a quarter. GDP is used as an indicator for government and economic decision makers for planning and policy making. Understanding what the future holds is also important for government officials, helping them to determine which fiscal and monetary policies to implement. Hence forecasting accuracy plays a vital role in GDP. A hybrid model has been considered an effective way to improve the forecast accuracy. In this paper, hybrid model of SARIMA-ELM is proposed for forecastingthe GDP of India to help policy making and plan future operating activities. Data from 1st quarter of 1999 to 1st quarter of 2020 is used for the analysis. Forecasting performance of hybrid model SARIMA-ELM is compared with traditional SARIMA model, ELM, MLP and SARIMA-MLP. Based on accuracy measures we concluded that the hybrid model of SARIMA-ELM is more appropriate model for forecasting the GDP. The best model is used to forecast GDP of India for next 2 years.

Keywords: GDP, Traditional time series model, ELM, MLP, SARIMA-ELM, SARIMA-MLP, Accuracy measures

1. INTRODUCTION:

GDP is one of the primary indicators used to measure the healthiness of a country's economy. It is the total market value of all final goods and services produced within a country in a specific time period- monthly, quarterly or annually. It includes private and public consumption, investment, and exports less imports. Policy makers, investors, economists, businesses, bankers, politicians, and even the media keep a close watch on GDP estimates. Gross Domestic Product is important in an economy because it is used to determine if an economy is growing more quickly or more slowly. Also, it is used to compare the relative growth rate of economies throughout the world. In India, The Central Statistics Office (CSO) under the Ministry of Statistics and Program Implementation is responsible for collecting and calculating Indian GDP data. The eight sectors considered for GDP calculation are: Agriculture, forestry, and fishing Mining and quarrying Manufacturing Electricity, gas, water supply and other utility services Construction Trade, hotels, transport, communication and services related to broadcasting Financial, real estate and professional services Public administration, defence and other services. Understanding what the future holds is also important for government and economic decision makers for planning and policy making. Hence this study have been taken up to forecasting future GDP of India which will help the Government entities use forecasts to plan their policy-making efforts as well. Fiscal policies and monetary policies are implemented based on the expectation of GDP growth. If the growth in GDP is expected to be strong, the government may enact tighter policies. On the other hand, if GDP growth is expected to be slow, the government may enact expansionary policies. Investors also use GDP growth forecasts to make informed decisions. If the economy is expected to be strong, they may be more comfortable investing in riskier assets, whereas if the economic conditions are expected to weaken, investors may be more conservative with their asset allocation.

2. METHODOLOGY

2.1. Introduction to Time series Analysis:

A Time Series is a sequence of observation ordered in equally spaced, discrete time intervals. A basic assumption in any time series analysis / modeling that some aspects of past pattern will continue to remain the future. Suitable forecasting time series model can be developed with minimum forecasting error. Atleast 50 observations are necessary for performing TS analysis, as propounded by Box Jenkins who were pioneers in TS modeling. The four main objectives in time series analysis are Description, Explanation, Prediction and Control. Time series analysis start by plotting the data and look for non-stationary components. Then eliminate these components using different methods, in order to have a stationary data. After identifying a suitable probability model for the time series, this model can be used for prediction. The statistical methodology used for analyzing time series is referred to as Time Series Analysis.

Seasonal ARIMA Model:

The seasonal autoregressive integrated moving average model of Box and Jenkins (1970) is given by,

$$\emptyset_p(B^s)\ \emptyset(B)(1-B^s)^D(1-B)^d X_t = \theta_p(B)\theta_Q(B^s)\varepsilon_t$$

The non- seasonal components are,

$$\emptyset(B) = 1 - \emptyset_1 B - \cdots - \emptyset_p B^P$$

$$\theta\ (B) = 1 + \theta 1 B + \cdots + \theta q B^q$$

The seasonal components are,

$$\Phi\ (B^s) = 1 - \Phi 1 B^s - \cdots - \Phi p B^{Ps}$$
$$\Theta\ (B^s) = 1 + \Theta 1 B^s - \cdots - \Theta q B^{Qs}$$

To find SARIMA fitted model we used The Box-Jenkins method (1970) and following steps:
1. Identification (Identification of the time series model that summarizes the data in the best possible way).
2. Estimation (Estimation of the parameters of the model identified in the previous step).
3. Diagnostic Checking (Evaluation of the model for better predictions about future).

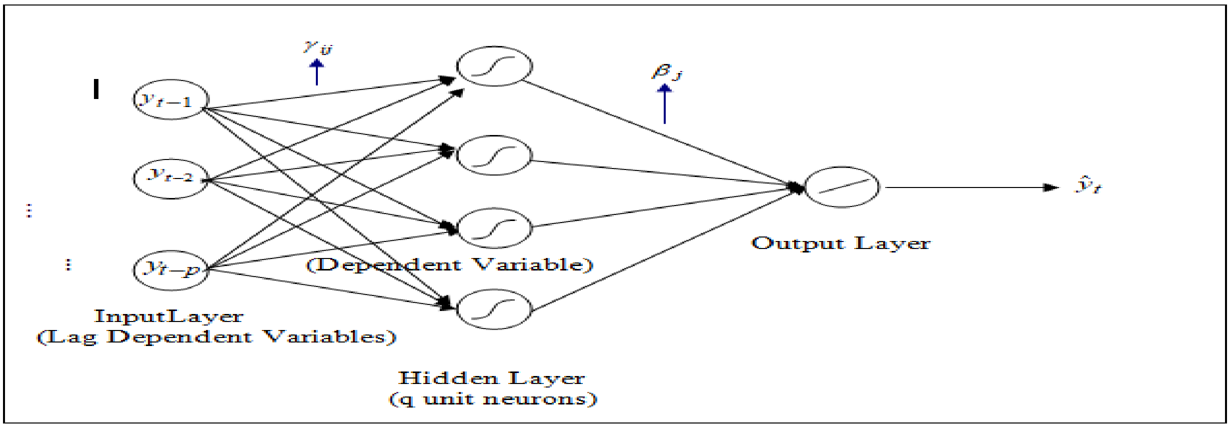
2.2 Multi Layer Perceptions – Neural Networks (MLP-NN):

A multilayer feed-forward neural network consists of an input layer, one or more hidden layers, and an output layer. A multilayer feed forward neural network is an interconnections of perceptron‘s in which data and calculations flow in a single directions, from the input data to the outputs. The number of layers in a neural network is the number of layers of perceptron‘s. There is no restriction on the number of hidden layers. The back propagation algorithm performs learning on a multilayer feed-forward neural network. It iteratively learns a set of weights for prediction of the class label of tuples. However, increases the number of perceptron‘s increases the number of weights that must be estimated in the network, which in turn increases the execution time for the network. Instead of increasing the number of perceptron‘s in the hidden layer to improve accuracy, it is sometimes better to add additional hidden layers, which typically reduce both the total number of network weights and the computational time.

MLP-NN model with a single hidden layer is represented as,

$$\boldsymbol{y_t} = \boldsymbol{\beta_0} + \sum_{j=1}^{q} \boldsymbol{\beta_1}\, \boldsymbol{g}\left(\Upsilon_{0j} + \sum_{i=1}^{p} \Upsilon_{ij}\boldsymbol{y_{t-i}}\right) + \boldsymbol{e_t}$$

Where, $\boldsymbol{y_{t-i}}$ $(i = 1,2, \dots, p)$ are the p inputs and $\widehat{\boldsymbol{y}_t}$ is the output; βj $(j = 0,1,2, \dots, q)$ and Υ_{ij} $(i = 0,1,2, \dots, p; j = 0,1,2, \dots q)$ are the connection weights and $\boldsymbol{e_t}$ is random error term, The integers p, q are the number of input and hidden nodes each; $\beta 0$ and γoj are the error terms and *g(.)* is activation function.



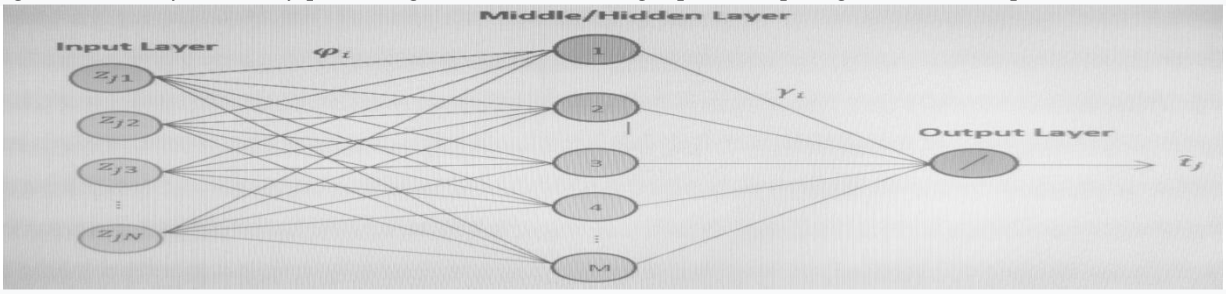
Architectures of MLP-NN with a single hidden layer

2.3. Extreme Learning Machine (ELM):

Extreme learning machine are feed-forward neural networks for classification, regression, clustering and prediction, compression and feature learning with a single layer or multiple layers of hidden nodes, where the parameters of hidden nodes need not be tuned. These hidden nodes can be randomly assigned and never updated or can be inherited from their ancestors without being changed. In most cases, the output weights of hidden nodes are usually learned in a single step, which essentially amounts to learning a linear model. In most cases, ELM is used as a single hidden layer feed-forward network. These models are able to produce good generalization performance and learn thousands of times faster than networks trained using back propagation.

Back propagation:

Back propagation is a method used in artificial neural networks to calculate a gradient that is needed in the calculation of the weights to be used in the network. Back propagation algorithm: Back propagation is a neural network learning algorithm. Back propagation learns by iteratively processing a data set of training tuples, comparing the network's prediction for each tuple with



the actual known target value. The target value may be the known class label of the training tuple (for classification problems) or a continuous value (for numeric prediction). For each training tuple, the weights are modified so as to minimize the mean-squared error between the network's prediction and the actual target value. These modifications are made in the ―backwards direction (i.e., from the output layer) through each hidden layer down to the first hidden layer. In general, the weights will eventually converge, and the learning process stops.

2.4 SARIMA-MLP and SARIMA-ELM hybrid model:

Hybrid model is a combination between linear and nonlinear models that usually be used for increasing the forecast accuracy. In general, the mathematical form of combination between linear and nonlinear models is as follows: $\mathbf{Z_t} = \boldsymbol{Y_t} + \boldsymbol{N_t}$ + et , Where $\boldsymbol{Y_t}$ is a linear component and Nt is a nonlinear component of the model. In this paper, MLP and ELM is used for modeling the nonlinear component separately.

Estimation of this hybrid model is done in two steps. The first is modeling the linear component to get the residual and then applying a nonlinear model to this residual for handling the nonlinear component. In this paper, SARIMA model is used for handling the linear component. Assume at is residual at period t from the SARIMA model.

i.e. $b_t =: \mathbf{Z_t} - \widehat{\boldsymbol{Y}}_t$, Where $\widehat{\boldsymbol{Y}}_t$ is the forecast of linear model at period t.

In Hybrid Model **SARIMA-MLP**,

MLP is applied for modeling b_t as follows:

$$b_t = \beta_0 + \sum_{j=1}^{q} \beta_1 g\left(\Upsilon_{0j} + \sum_{i=1}^{p} \Upsilon_{ij} b_{t-i} \right) + \epsilon_t = \widehat{N}_t + e_t$$

Where the first two part is the non-linear function of MLP and e_t is the residual of this MLP model. Hence, the forecast value of the hybrid SARIMA-MLP model is as follows,

$$\widehat{Z}_t = \widehat{Y}_t + \widehat{N}_t.$$

Similarly, In hybrid **SARIMA-ELM** model, for the residual seriesb_t, Constructed the nonlinear model and the forecast value of the hybrid SARIMA-ELM model is as follows,

$$\widehat{Z}_t = \widehat{Y}_t + \widehat{N}_t$$

2.5 Accuracy Measures

In forecasting, our objective is to produce and optimum forecast that has no error or as little error as possible, which leads us to the minimum mean square error forecast. This forecast will produce an optimum future value with the minimum error in terms of the mean square error criterion.

$$\text{The Root Mean Square Error (RMSE)} = \sqrt{\frac{\sum_{t=1}^{n}(Y_t - \widehat{Y}_t)^2}{n}}$$

Mean Absolute Error (MAE) = $\sum_{t=1}^{n} \frac{|Y_t - \widehat{Y}_t|}{n}$

$$\text{Mean Absolute Percentage Error (MAPE)} = \frac{1}{n}\sum_{t=1}^{n} \frac{(Y_t - \widehat{Y}_t)}{Y_t} \text{ X } 100$$

Where Y_t = the actual value, $\widehat{Y}_t$ = the forecasted value, and n = the size of test set.

3. ANALYSIS AND DISCUSSION

This data set used in the study is download from the RBI website. We collected GDP data from 1st Quarter of 1999 to the 1st Quarter of 2020 .Where Q_1, Q_2, Q_3 &Q_4 denotes; April to June, July to September, October to December and January to March quarters, respectively. The data from 1999 to 3rd Quarter of 2017 is used for model fitting. The rest of data is used as test data to check the accuracy of forecast. The analysis is carried out using R software.

In this section, five different models such as, traditional SARIMA model, ELM, MLP and SARIMA-MLP and SARIMA-ELM are fitted for the data and accuracy of fitted models are presented, Where SRIMA model is classic linear model ,ELM and MLP are modern nonlinear model, and hybrid model SARIMA-MLP and SARIMA-ELM that combining classical linear and modern nonlinear models.

Fig 1: *Time series plot for Gross Domestic Product*

From above figure we observe that data has both seasonal component and trend component.

Man Kendall test – test for trend

H0: There is no trend in the series v/s H1: There is a trend in the series.

tau = 0.929, 2-sided p-value = 2.22e-16 Since the computed p-value is less than the significance level alpha=0.05, we reject the null hypothesis and conclude that there is trend in the given data.

Rank-sum test –test for seasonality:

Ho: There is no seasonal variation in the data.

H_1: There is seasonal variation in the data.

Chi-square calculated value is 75.42353 and Chi-square critical value is 7.814728. Since calculated value of chi square is greater than critical value of chi -square we reject H0 and conclude that there is seasonal variation in the data. Then we removed seasonal component from the data.

Variance difference method:

Now data has only trend component. Therefore we carried out variance difference method to make the given series stationary. Variance of given time series is 31126160019.Variance of first difference series is 16073827437.Variance of second difference series is 34917113766. Since variance of second difference series is more than the variance of first difference series, first difference series is stationary.

Augmented Dickey-Fuller test:

H0: The series is not stationary

H1: The series is stationary

Dickey-Fuller = -5.1092, Lag order = 4, p-value = 0.01

If p value is less than 0.05, we reject H0 and conclude that the series is stationary.

SARIMA Model

Based on the ACF and PACF plot we fitted the different sarima model with different order. We select the best model for which AIC value is minimum and p value maximum.

Table 1: Model fitting

(p,d,q) (P,D,Q)	AIC	Box-pierce p-value	Ljung-Box p-value
(0,1,0) (0,1,1)	1805.386	0.99996	0.9885926
(0,1,0) (1,1,0)	1827.083	0.94458	0.4780385
(0,1,0) (1,1,1)	1807.145	0.979993	0.9682564
(0,1,0) (2,1,0)	1819.055	0.896312	0.7779937
(0,1,0) (2,1,1)	1808.145	0.989991	0.975986
(0,1,0) (3,1,0)	1805.419	0.979010	0.965758
(0,1,0) (3,1,1)	1810.645	0.986580	0.977785

From the Table 1, the AIC value for model SARIMA $(0,1,0)(0,1,1)_4$ is minimum and it has highest p-value . Thus SARIMA (0,1,0) $(0,1,1)_4$ is the best fitted model.

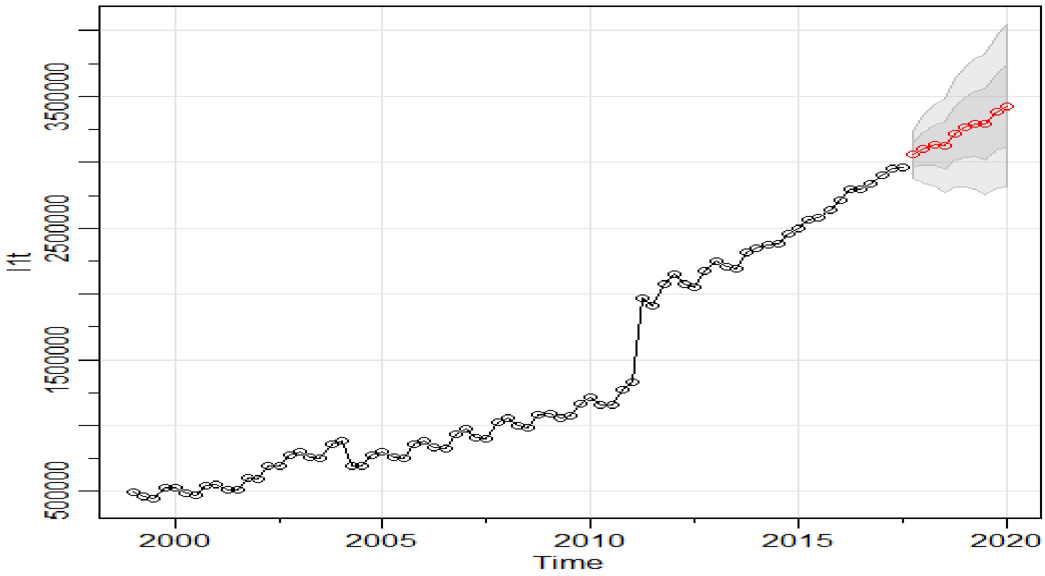


Fig.2: *Forecast from SARIMA (0,1,0) $(0, 1, 1)_4$ model*

Forecasting from neural network: MLP:

MLP fit with 5 hidden nodes and 20 repetitions. Deterministic seasonal dummies included. Forecast combined using the median operator

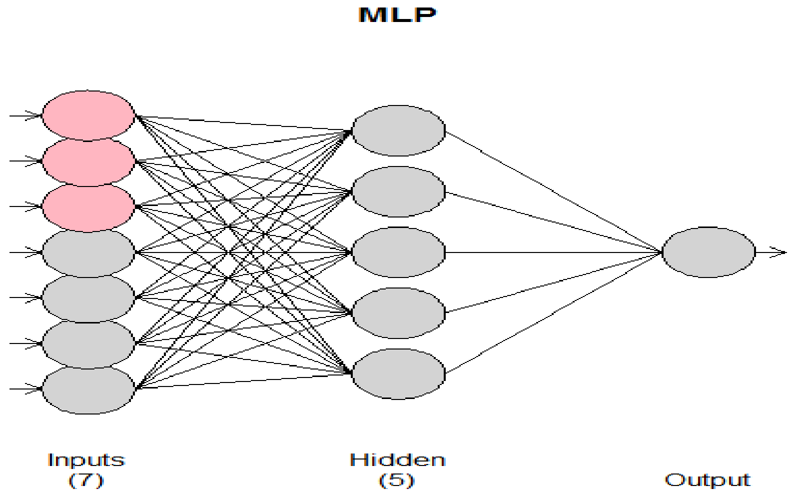


Fig.3: *MLP network*

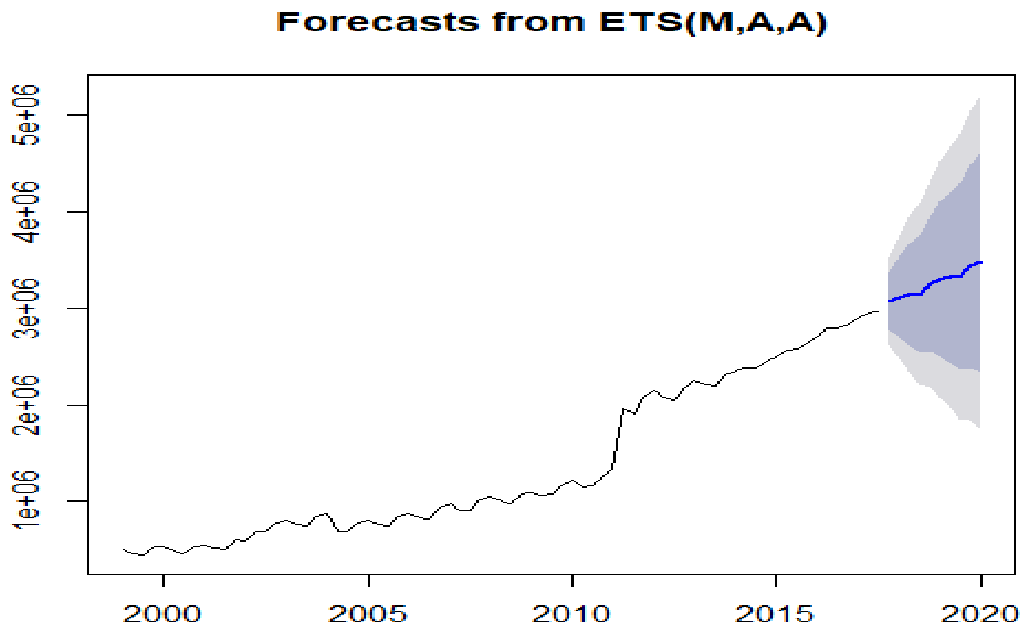


Fig.4: *Forecast from neural network.*

ELM:

ELM fit with 98 hidden nodes and 20 repetitions. Deterministic seasonal dummies included. Forecast combined using the median operator. Output weight estimation using: lasso.

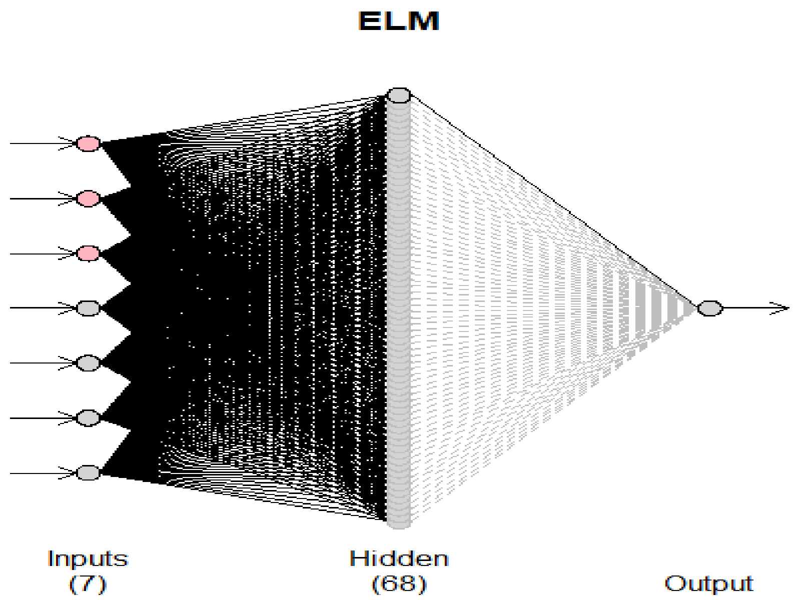


Fig.5: *ELM network*

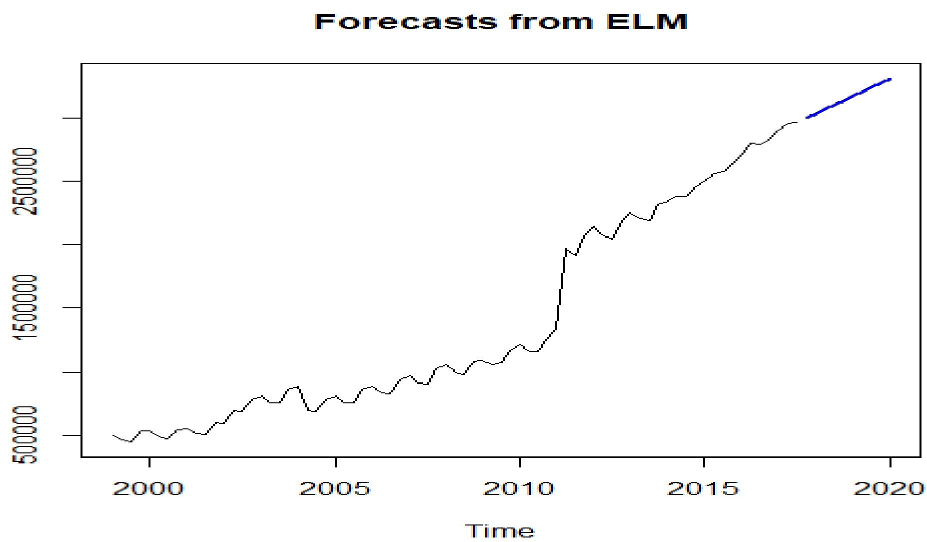


Fig.6: *Forecast from Elm network*

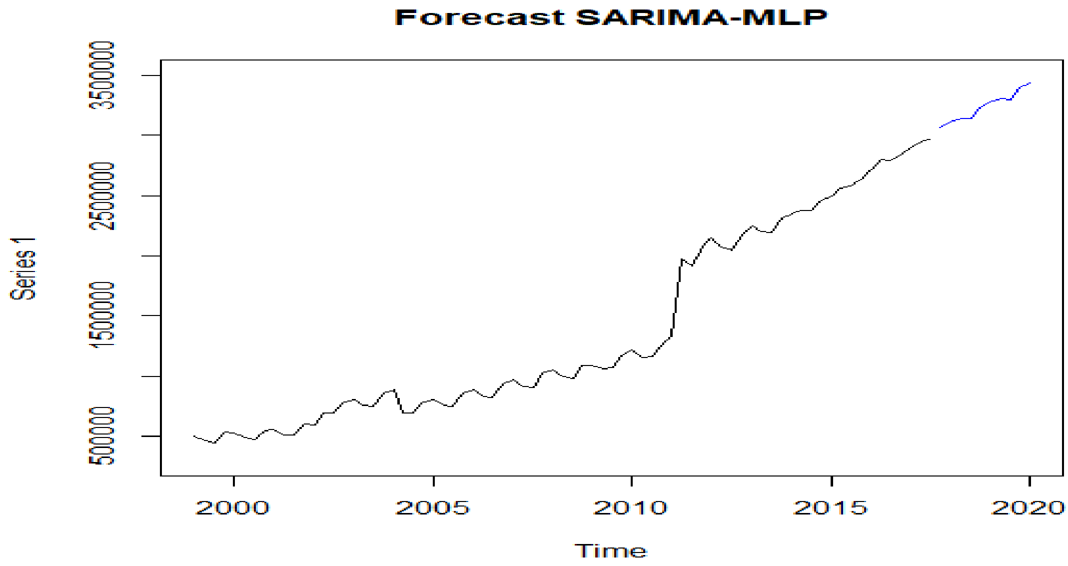


Fig.7: *Forecast from SARIMA-MLP hybrid model.*

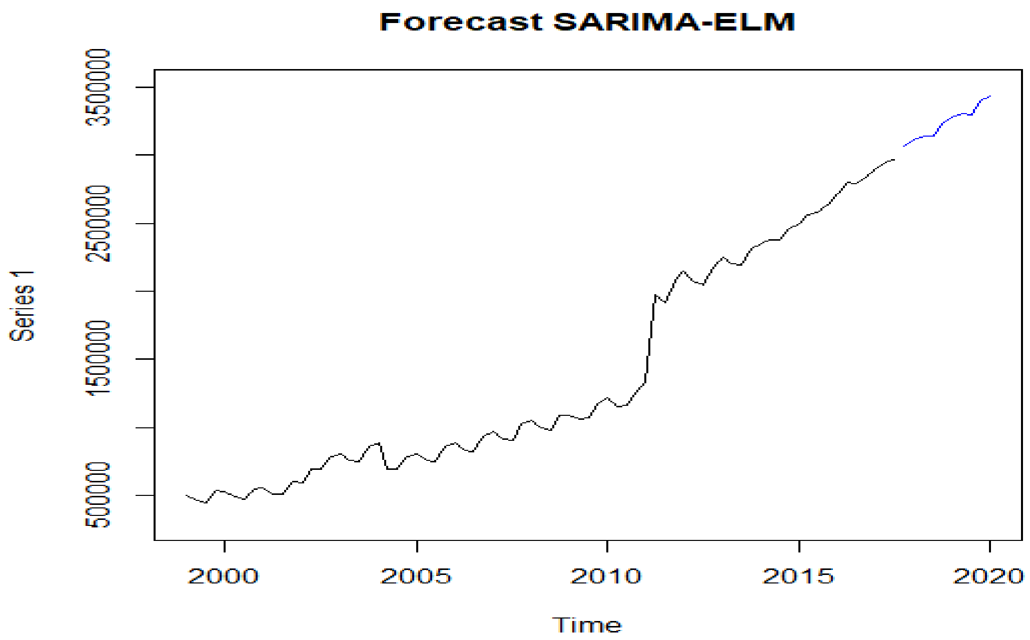


Fig.8: *Forecast from SARIMA-ELM hybrid model.*

Table 2:. Comparison between actual value and forecasted value of all models

Actual values	SARIMA	MLP	ELM	SARIMA-MLP	SARIMA-ELM
3036177	3058290	3073917	2997326	3071165	3071279
3123451	3102333	3116771	3032060	3115208	3115321
3157366	3131235	3141062	3096794	3144109	3144223
3142027	3125206	3143907	3111528	3138080	3138194
3206925	3220904	3255696	3136262	3233779	3233892
3296807	3264947	3298551	3190996	3277822	3297935
3307707	3293849	3322842	3255730	3306723	3306837
3278144	3287820	3325687	3240464	3300694	3300808
3318220	3383518	3437476	3281598	3396393	3316506
3397049	3427561	3480330	3309932	3440435	3410549

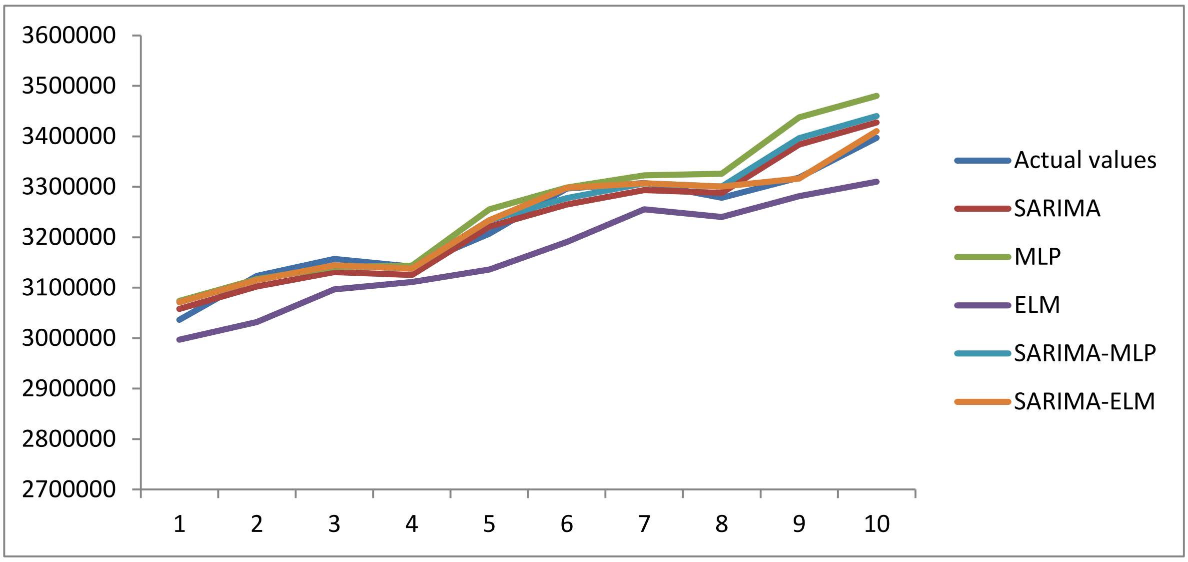


Fig 9: *Plot of actual and forecasted value of all models*

Table 3: Accuracy measure for the models SARIMA(0,1,0)(0,1,1)$_4$, MLP ,ELM and hybrid models:

	Models	RMSE	MAE	MAPE
1	SARIMA	29311.62	25136.67	0.77499
2	MLP	94850.88	75823.93	2.355393
3	ELM	37799.14	29428.63	1.128548
4	SARIMA-MLP	31268.34	25134.67	0.772567
5	SARIMA-ELM	26323.29	25036.67	0.770546

From the above table we observe that accuracy measures RMSE,MAE and MAPE are low for SARIMA-ELM hybrid model. Therefore SARIMA-ELM hybrid model is the best model for forecasting the GDP .Then we forecast the GDP for year next 2 years based on hybrid model. The forecasted graph is given below:

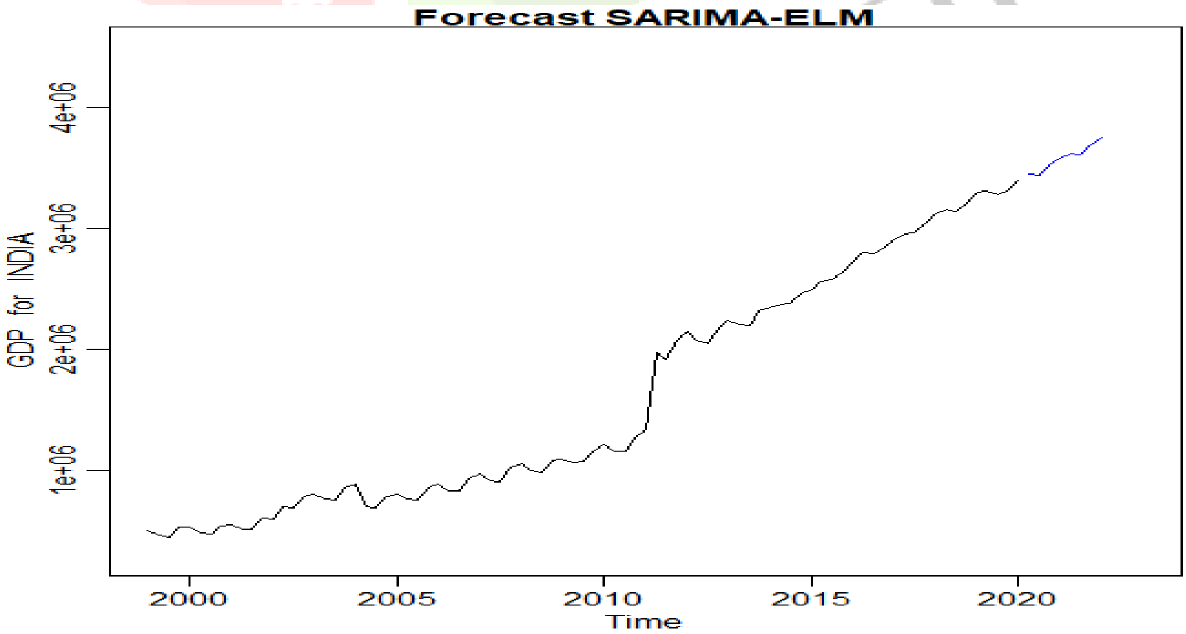


Fig.10*: Future Forecast from SARIMA-ELM hybrid model*

4. CONCLUSION

GDP refers to the total market value of all goods and services that are produced within a country per quarterly or yearly. It is an important indicator of the economic strength of a country. Understanding what the future holds is also important for government and economic decision makers for planning and policy making. This paper presents five different models namely SARIMA, MLP, ELM, SARIMA- MLP and SARIMA-ELM. We compared the forecasting accuracy of different models based on MAPE, MAE and RMSE. Based on table 3, the forecasts produced by SARIMA – ELM are better since the RMSE, MAE, MAPE are lower than the forecasts produced by other models. It can be concluded that in case of GDP for India, hybrid model improves the forecasting accuracy and can be used as the best model for predicting future GDP. Finally, we forecasted the GDP of India using the best model for the 2 years.

References

1. Box GEP and Jenkins G.M (1976): Time Series Analysis: Forecasting and Control, Holden-day, San Franscisco.
2. Brockwell P.J. and Davis R.S. (2002): Introduction to Time Series and Forecasting, 2nd Ed., Springer.
3. Deelip kumar shetty ,Sumithra and Ismail B (2018): Hybrid SARIMA-GARCH model for forecasting Indian Gold price ,Research Review International Journal of Multidisciplinary,Volume-03 ,Issue-08
4. Jiawei Han, Micheline Kamber (2002): Data Mining-Concepts and Techniques, Morgan Kaufman Publishers, U.S.A
5. Montgomory D.C. and Johnson L.A. (1977): Forecasting and Time Series Analysis, McGraw Hill.
6. Sandhya Samarasinghe (2007): Neural Networks for Applied Sciences and Engineering, Auerbach Publications, Taylor and Francis Group, New York.
7. Satyanarayana ,Swathi and Ismail B : Statistical Analysis of Food Grain Prices in Karnataka, Research Review International Journal of Multidisciplinary, Volume-03 Issue-08
8. Simon Haykin (2012): Neural Networks and Learning Machine, 3rd edition, PHI Learning Private Limited ,New Delhi.
9. Sunil, Satyanarayana, Sachin Acharya, Arun Kumar Jogi :Application Of Hybrid Model For Forecasting Prices Of Jasmine Flower In Bangalore, India , International Journal Of Scientific & Technology Research Volume 8, Issue 11, November 2019
10. Waciko KJ, Ismail B. SARIMA-ELM Hybrid Model for Forecasting Tourist in Nepal. RESEARCH REVIEW International Journal of Multidisciplinary. 2018; 3(7):343-349.