

# Stock Market Prediction using Auto Regressive Integrated Moving Averages (ARIMA)

K Murali<sup>1</sup>, Sk Nafeez Umar<sup>2</sup>, D Chandrakesavulu Naidu<sup>3</sup>, MP Reddeppa Reddy<sup>4</sup>, C Mani<sup>5</sup>,  
B Ramana Murthy<sup>6</sup>

<sup>1</sup>Academic Consultant, Dept. of Statistics, S V University, Tirupati, India

<sup>2</sup>Assistant Professor, S V Agricultural College, Tirupati, India

<sup>3</sup>Associate Professor, Dept. of Statistics, S G S Arts College, Tirupati, India

<sup>4</sup>Associate Professor, Dept. of Statistics, S V Arts College, Tirupati, India

<sup>5</sup>Principal, S G S Arts College, Tirupati, India

<sup>6</sup>Assistant Professor, S V Agricultural College, Tirupati, India

**Abstract:** Stock market prediction plays an important role to decide investment in markets over the time period. The Auto Regressive Integrated Moving Averages (ARIMA) have been explored for time series prediction. This paper explores the process and method of building Stock predictive models is using ARIMA model. The Stock market indices data of Bombay Stock Exchange (BSE) is used in building stock predictive model. The Results revealed that the ARIMA model has a robust for particularly short term prediction and endorsed with current techniques for Stock Market prediction. The study made a few observations which may help the investors and model builders to understand better about the stock market analysis.

**Keywords:** Stock market prediction, ARIMA, Short term prediction.

## 1. Introduction

Karamouz and Araghinejad (2012), ARIMA models are non-static and cannot be used to reconstruct the missing data. Balaguer *et al.* (2008), time series models (ARMA and ARIMA) and artificial neural networks in different fields of hydrology. Toth *et al.* (2000), used the artificial neural network and ARMA models to forecast rainfall. Mohammadi *et al.* (2005), by using artificial neural network and ARMA models and regression analysis forecasted Karaj reservoir inflow using data of snow melting. According to Chegini (2012), the highest inflows were in the spring due to the snow melt because of draining in threshold winter.

Several researchers worked on ARIMA forecasting models to predict the future returns (M. Khasel *et al.*, 2009; C. Lee & C. Ho, 2011; M. Khashei *et al.*, 2012). Itimi *et al.*, (2018) conclusion on the result of the theory of a positive risk premium on stock indices which states that the higher returns are expected for asset with higher level of risk. This confirms that the Nigerian stock market prices are reactive to the changes in macroeconomic variables eventually regardless of high volatility and immaturity.

ONUOHA (2018) insights in the predictability of G.C.C stock returns using crude oil prices using the approach of Wasteland *et al.* (2012, 2015) reveal of salient features of the predictor. The results show superior performance of the oil based stock model over time-series models (namely, AR, MA, ARMA and ARIMA) for both in-sample and out-of-sample forecasts. The results are robust to different oil prices series (Brent and WTI prices) and forecast horizons (30 and 60 days). According Mohammed Nafie, Alkesh (2019), the

Sultanate economy is projected to achieve a positive growth of 3 percent at constant prices. The International Monetary Fund (IMF) expects OMAN to have the highest growth rate among all G.C.C countries in 2019, which is reflective of positive potential opportunities for local and international investments.

## 2. Methodology

The current study is based on the weekly closing market index for the Bombay Stock Exchange (BSE). Weekly data of January 2010 - December 2020 is used for forecasting models. In this study training, testing and complete data set are used for model building. Firstly 60% of data is used for training, 40% of data is used for testing using ARIMA models. The Statistical software's R-Language and GRETLE were used for the forecasting model building.

### a) Test for Normality

All assumption were used in the study of model building. InJarque-Bera normality test was used in this analysis for normality of the data. The null hypothesis for this is data follows normal distributed; the alternative hypothesis is the data does not follow normality. The data uses normalise technique for time series indices of BSE indices data

$$z = \frac{x - \min(x)}{\max(x) - \min(x)}$$

### b) Augmented Dickey-Fuller (ADF) Test:

One of the imperative assumptions on time series is non-stationary relationship. To make sure existence of stationary relationship, the following stationary test and Augmented

Dickey–Fuller test (ADF) were employed in the study.

$$\Delta\lambda_t = \alpha_0 + \alpha_2 t + \sum_{i=1}^k \beta \Delta\lambda_{t-1} + \varepsilon_t$$

Where  $\lambda_t$  denotes the weekly index of the individual stock at time t,  $\beta$  is the coefficient to be estimated, k is the number of lagged terms, t is the trend term,  $\alpha_2$  is the estimated coefficient for the trend,  $\alpha_0$  is the constant and  $\varepsilon$  is white noise.

**c) Performance of the Model:**

Literature revealed the use of measure to evaluate forecast accuracy such as scale dependent, percentage error and relative error a study of forecast accuracy. For evaluating the performance of the model Mean Absolute Percentage Error (MAPE) has been used which is defined as follows.

$$MAPE = \frac{1}{n} \sum_{t=1}^n \left| \frac{A_t - F_t}{A_t} \right| \times 100$$

Where  $A_t$  is the actual value and  $F_t$  is the forecast value.

**d) The Box-Jenken’s Methodology**

The ARIMA (p d q) model is churn of autoregressive (AR) and moving average (MA) model which shows that there is relation between present value and past value and residuals respectively. Box-Jenken’s methodology applied for time series data identified best ARIMA models and residual. The time series data follows the Box-Jenkins (1970) methodology for modeling, generally known as ARIMA. Let  $Y_t$  be a discrete time series variable which takes different variable over a period of time. The corresponding AR (p) model of series, which is the generalization of the autoregressive model, is expressed as follows.

$$Y_t = \theta_1 Y_{t-1} + \theta_2 Y_{t-2} + \dots + \theta_p Y_{t-p} - \phi_1 \varepsilon_{t-1} - \phi_2 \varepsilon_{t-2} \dots - \phi_q \varepsilon_{t-q} + \varepsilon_t$$

**3. Results and Discussion**

Previous 582 observations were used for this model. Numerous methods were studied to judge the suitable model. The data was divided into 60% and 40% of total BSE indices. The first 60% of data is consider as training model and remaining 40% data is consider as testing model. The suitable model has been selected based on the maximum coefficient of determination, minimum Mean Absolute Percentage Error (MAPE) and minimum Bayesian information criterion (BIC) respectively

**Table 1:** Results of Unit Root Test (Augmented Dickey Fuller Test)

	Constant/Trend	ADF test	p-value
Training model	Constant and trend	-1.7088	0.6997
Testing model	Constant and trend	-1.9208	0.6094
Overall model	Constant and trend	-2.407	0.406

Alternative hypothesis: Stationary

The Jarque-Bera were used for normality test for BSE indices data and the result of Jarque-Bera test is 34.034 which is significant at 0.01 level. The results of Augmented Dickey Fuller (Unit Root test) for BSE indices of training,

testing and overall model data are not significant with respective probability values, which is stationary at first order of integration level where as not significant at 0.05 level.

**Table 2:** Residual analysis of BSE indices

Model	R-square	MAPE	BIC
Training model	0.9851	1.922	12.632
Testing model	0.9500	1.991	14.091
Overall model	0.9930	1.864	13.251

Various mixtures of ARIMA(P D Q) (p d q) models were used in this study and appropriate models were selected based on selection of criteria. Among the selected models ARIMA (2 1 2)(2 1 2) model was found most appropriate and best model for predicting the BSE indices in future. Of the 349 observations that were considered for training period, the R- square value was 0.9851, MAPE was only1.922, which indicates accuracy of the model. In the testing period 233 observation were considered and observed that R-square value was 0.95 and MAPE was 1.991. The overall model also showed that R-square value is 0.9930 and MAPE was 1.864 which shows accuracy of the model. When we compare all the stages of training, testing and overall model, the same as an R-square value, MAPE and BIC values. Which shows the ARIMA gives the best model for times series data. (Table 2)

The method and approach of the predicting model used and compared training, testing and overall models. All the three procedures observed very close to R-square value and Bayesian Information Criteria. The results reveal that all the three models were fit in to ARIMA (2 1 2) (2 1 2) model. The trend of BSE indices values shows positively increasingly in future forecasting (Figure3) and the residuals were shows consistently mean and variance (Figure5).

ARIMA (2 1 2)(2 1 2) model is used for predicting BSE Indices for upcoming months of January 2021 to December 2021. Steadily performing future indices from January 2021 to December 2021 are with 95% confidence interval.

**Table 3:** Predicted values of BSE Sensex (January 2021 to December 2021)

Months	Predicted	95% confidence Limits	
		Lower	Upper
Jan-21	51653.3	50256.9	53049.8
Feb-21	51802.2	49820.7	53783.7
Mar-21	52152.4	49655.9	54648.8
Apr-21	52157.4	49223.6	55091.1
May-21	52194.2	48886.8	55501.7
Jun-21	52665.2	49024.9	56305.5
Jul-21	52824.8	48879.3	56770.3
Aug-21	53153.9	48924.8	57383.1
Sep-21	52809.2	48314.2	57304.2
Oct-21	52815.3	48069.4	57561.2
Nov-21	52994.3	48010.1	57978.5
Dec-21	52721.1	47509.5	57932.7

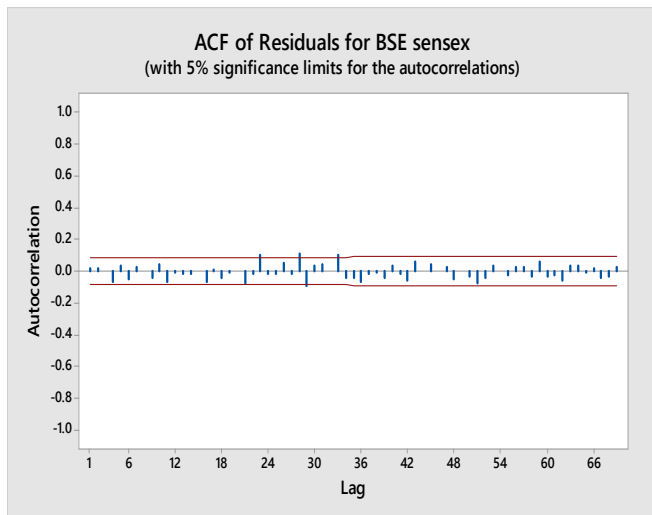


Figure 1: ACF of residual for overall data

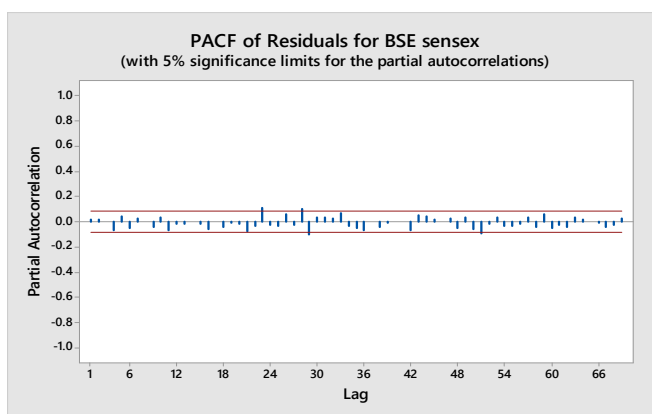


Figure 2: PACF of residual for overall data



Figure 3: Predicted trend for BSE indices

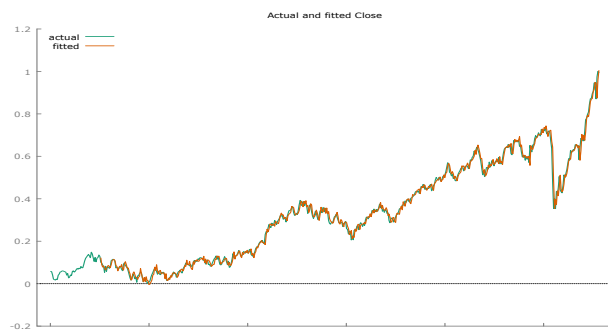


Figure 4: Trend analysis of Actual and Predicted

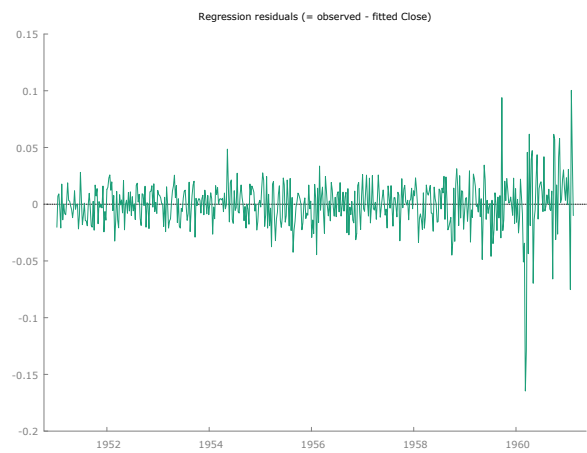


Figure 5: Residual analysis of BSE indices

#### 4. Conclusion

The objective of this study was to predict BSE indices in India. The Auto Regressive Integrated Moving Average (ARIMA) model was used for this objective. Time series data and closing index for 581 weekly observations have been considered in the model. Data divided in to two phases firstly training data set and secondly testing data set. Box Jenkin’s methodology is used for predicting weekly indices. Identified the appropriate ARIMA (2 1 2) (2 1 2) model of BSE indices. The predicted values are useful to invest in markets as well as to predict pattern of BSE indices.

#### References

- [1] Karamouz M and AraghinejadSh 2012 Advance Hydrology. Amirkabir University of Technology Press.
- [2] Balaguer E, Palomares A, Sorie E, Martin- Guerrero J D 2008 Predicting service requestin support centers based on nonlinear dynamics, ARMA modeling and neural networks. Expert Syst. Appl. 34, 665–672.
- [3] Toth E, Brath A and Montanari A 2000 Comparison of short-term rainfall predication models for real-time flood forecasting. J. Hydrol. 239, 132–147.
- [4] Mohammadi K, Eslami H R and Dayyani DardashtiSh 2005 Comparison of regression ARIMA and ANN models for reservoir inflow forecasting.
- [5] Chegini A G 2012 MATLAB Tools, Naghous Press. <http://www.naghooospress.ir/bookview.aspx?bookid=1485875>.
- [6] Onuoha I. N. (2018). “Improving the forecast accuracy of oil-stock nexus in GCCcountries”, *Theoretical Economic Letters*, Vol. 8, pp. 3073-3091, doi: https://doi.org/10.4236/tel.2018.814191.
- [7] Mohammed Nafie and Alkesh Joshi (2019). “Oman budget 2019: Walking a tightrope”, available online at: https://www.businessliveme.com/oman-budget-2019-walking-a-tightrope/10/02/2019.
- [8] Fama E. F. (1965). “Random walks in stock market prices”, *Financial Analyst Journal*, Vol. 21, No. 5, pp. 55-59, doi.10.2469/faj.v51.n1.1861.
- [9] Shaik Nafeez Umar, Labeeb Mohammed Zeeshan (2019), Modelling Monthly Volatility of the Muscat Securities Market (MSM) Index Using Auto Regressive Integrated Moving Average (ARIMA),

Journal of Business and Economics, ISSN 2155-7950,  
USA, November 2019, Volume 10, No. 11, pp. 1045-  
1056

- [10] Ramesh D, Banjul Bhattacharyya, Ria Biswas (2014) Forecasting of Maize Production in Andhra Pradesh by ARIMA Modeling, Environment and Ecology 32 (4B): 1709-1713.
- [11] Sivaramane N. Mathur V (2010) Forecasting of rice exports from India: An application of Box-Jenkins methodology. Agric Situation India 67 321.325