

# Indian Share Market Forecasting with ARIMA Model

Swapnil Jadhav<sup>1</sup>, Saurabh Kakade<sup>2</sup>, Kaivalya Utpat<sup>3</sup>, Harshal Deshpande<sup>4</sup>

Department of Computer Engineering, Savitribai Phule Pune University<sup>1,2,3,4</sup>

**Abstract:** Artificial neural networks (ANNs) are a flexible computing frameworks and universal approximates that can be applied to a wide range of time series forecasting problems with a high degree of accuracy for the convenience of predicting the futuristic value in share market and give a better future scope for investment. But yet the artificial neural network is not to the satisfactory as it includes both theoretical and empirical findings have concluded that the combination of different models can be an effective way of improving upon the predictive performance, if the models in ensemble are quite different. In this paper, a novel hybrid model of artificial neural network is proposed using a auto-regressive integrated moving average (ARIMA) model to produce the more accurate forecasting model than artificial neural network. On this context, we collected data on monthly closing stock indices of sensx, on these we have tried to develop an appropriate model that would help us to forecast the future unknown values of Indian stock market indices, i.e, ARIMA. Therefore, it can be used as an appropriate alternative model for forecasting task, especially when higher forecasting accuracy is needed.

**Keywords:** Sensx, Time Series, ARIMA model, validation.

## I. INTRODUCTION

Share market is dynamic in nature means to predict share price is very complex process so use of neural network (ARIMA model) is a choice of interest for share market prediction. The main reason behind the proposed system is that there is no linear relationship between market parameters and target closing price. The concept of forecasting stock market return has become fairly popular maybe because of the fact that if the future market value of the stocks is successfully predicted, the investors may be better guided. The profitability of investing and trading in the stock market to a large extent depends on the predictability of the system which in turn prepares the investors in their encounter with their future insecurities and risks associated with the market.[1]

## II. LITERATURE REVIEW

Stock market is basically non-linear in nature, prediction of stock market plays important role in stock business. Data mining and neural network can be effectively used to uncover non-linearity of stock market.

Stock market forecasters focus on developing successful approach for forecast/predict index value of prices, Ultimate aiming of high profit using well-defined trading strategies. The vital idea to successful stock market prediction is achieving best result, also minimize inaccurate forecast stock price.

Several computing technique need to be combined in order to predict the nature of stock market. As the time elapsed, traditional capital market theory has been changed and various methods of financial analysis have been improved. Data mining can solve all the tasks of retrieving data such as mathematical figures and text documents, spatial and multimedia data and hypertext document, etc. For the past few years forecasting of stock return has been important field of research.[1-9]

## III. OBJECTIVE AND STUDY

In this study our objective is to forecast the future stock market indices using time-series ARIMA model.

## IV. DATA AND METHODOLOGY

Our analysis involves monthly data on the closing stock indices of sensx for three consecutive months based on which we have tried to establish a suitable probability model namely the ARIMA model to forecast the future unobserved indices of sensx. After obtaining the required data, our first task was to check whether it is suitable for our purpose or not.

ARIMA processes are a class of stochastic processes used to analyze time series. The application of the ARIMA methodology for the study of time series analysis is due to Box and Jenkins .In this section, the description of the proposed ARIMA model and the general statistical methodology are presented. The general scheme is as follows:

Step 0) A class of models is formulated assuming certain hypotheses.

Step 1) A model is identified for the observed data.

Step 2) The model parameters are esti-mated.

Step 3) If the hypotheses of the model are validated, go to Step 4 otherwise go to Step 1 to refine the model.

Step 4) The model is ready for fore-casting.

### A. Step 0

In this step, a general ARIMA formulation is selected to model the price data. This selection is carried out by careful inspection of the main characteristics of the hourly price series. In most of the competitive electricity markets this series presents: high frequency, nonconstant mean and variance, and multiple seasonality (corresponding to daily and weekly periodicity, respectively), among others. If

denotes the electricity price at time  $t$ , the proposed general ARIMA formulation is the following:

- price at time
- functions of the backshift operator
- the error term.

Functions have special forms. They can contain factors of polynomial functions of the form where several values of  $\alpha$  and  $\beta$  can be set to 0.

It should be noted that this example does not correspond to a standard ARIMA formulation, as presented in [1]. However, the model is sufficiently general to include the main features of the price data. For example, to include multiple seasonality, factors of the form  $\alpha(B^k)$  can be included in the model. Finally, certain hypotheses on the model must be assumed. These hypotheses are imposed on the error term,  $\epsilon_t$ .

In Step 0, this term is assumed to be a randomly drawn series from a normal distribution with zero mean and constant variance that is, a white noise process. In Step 3, a diagnosis check is used to validate these model assumptions.

#### B. Step 1

A trial model, as seen in (1), must be identified for the price data. First, in order to make the underlying process stationary (a more homogeneous mean and variance), a transformation of the original price data and the inclusion of factors of the form  $\alpha(B^k)$  may be necessary. In this step, a logarithmic transformation is usually applied to the price data to attain a more stable variance. And, to attain a more stable mean, factors of the form  $\alpha(B^k)$  may be necessary, depending on the particular type of electricity market, as explained at the end of this section. After the underlying process is accepted as being stationary, the structure of functions and in (1) must be selected. In a first trial, the observation of the autocorrelation and partial autocorrelation plots of the price data can help to make this selection. In successive trials, the observation of the residuals obtained in Step 3 (observed values minus predicted values) can help to refine the structure of the functions in the model.

#### C. Step 2

After the functions of the model have been specified, the parameters of these functions must be estimated. Good estimators of the parameters can be computed by assuming the data are observations of a stationary time series (Step 1), and by maximizing the likelihood with respect to the parameters  $\theta$ .

The SCA System is used to estimate the parameters of the model in the previous step. The parameter estimation is based on maximizing a likelihood function for the available data. A conditional likelihood function is selected in order to get a good starting point to obtain an exact likelihood function, so, an option to detect and adjust possible unusual observations (called outliers in the forecasting literature) is selected. As these events are not initially known, a procedure that detects and minimizes the effect of the outliers is necessary. With this adjustment, a

better understanding of the series, a better modeling and estimation, and, finally, a better forecasting performance is achieved.

#### D. Step 3

In this step, a diagnosis check is used to validate the model assumptions of Step 0. This diagnosis checks if the hypotheses made on the residuals (actual prices minus fitted prices, as estimated in Step 1) are true. Residuals must satisfy the requirements of a white noise process: zero mean, constant variance, uncorrelated process and normal distribution. These requirements can be checked by taking tests for randomness, such as the one based on the Ljung-Box statistic, and observing plots, such as the autocorrelation and partial autocorrelation plots. If the hypotheses on the residuals are validated by tests and plots, then, the model can be used to forecast prices. Otherwise, the residuals contain a certain structure that should be studied to refine the model in Step 1. This analysis is based on a careful inspection of the autocorrelation and partial autocorrelation plots of the residuals (see Appendix A).

#### E. Step 4

In Step 4, the model from Step 2 can be used to predict future values of prices (typically 24 hours ahead). Due to this requirement, difficulties may arise because predictions can be less certain as the forecast lead time becomes larger.

The SCA System is again used to compute the 24-hour forecast. Likewise, the exact likelihood function option and the detection and adjustment of outliers procedures are selected.

Note that, as mentioned in Step 0, the proposed formulation extends the standard ARIMA model by including more than two factors in  $\alpha(B^k)$  and a special polynomial structure of the overall function. It also should be noted that model needs the previous 5 hours to predict the next hour, whereas just needs the previous two hours. Also, the model in [1] does not use differentiation, and the one in [2] uses hourly, daily and weekly differentiation:  $\alpha(B^k)$ . This is related to the stationary property of the series, and it can be traced by inspecting the autocorrelation and partial autocorrelation plots.[5]

### V. ARIMA MODELLING

ARIMA models are the most general class of models for forecasting a time series which can be stationarized by transformations such as differencing and logging. The acronym ARIMA stands for "Auto-Regressive Integrated Moving Average." Lags of the differenced series appearing in the forecasting equation are called "autoregressive" terms, lags of the forecast errors are called "moving average" terms, and a time series which needs to be differenced to be made stationary is said to be an "integrated" version of a stationary series. Random-walk and random-trend models, autoregressive models, and exponential smoothing models (i.e., exponential weighted moving averages) are all special cases of ARIMA models. A non-seasonal ARIMA model is classified as an "ARIMA (p,d,q)" model, where:

- p is the number of auto-regressive terms,
- d is the number of non-seasonal differences,
- q is the number of lagged forecast errors in the prediction equation.

Generally a non-seasonal stationary time-series can be modeled as a combination of the past values and the errors which can be denoted as ARIMA (p, d, q) or can be expressed as:  $X_t = \phi_0 + \phi_1 X_{t-1} + \phi_2 X_{t-2} + \dots + \phi_p X_{t-p} + \epsilon_t - \theta_1 \epsilon_{t-1} - \theta_2 \epsilon_{t-2} - \dots - \theta_q \epsilon_{t-q}$ , where  $X_t$  and  $\epsilon_t$  are the actual value and random error at time t, respectively;  $\phi_i$  ( $i=1, 2, \dots, p$ ) and  $\theta_j$  ( $j=1, 2, \dots, q$ ) are model parameters. p and q are integers and often referred to as orders of autoregressive and moving average polynomials.

In order to evaluate the prediction performance, it is necessary to introduce a forecasting evaluation criterion. In this study, the quantitative evaluation is used as the accuracy measures.[1]

## VI. CONCLUSION

Multivariate time-series forecasting of Sensex . Forecasting accuracy will be enhanced if we study the probability distribution nature of Sensex. The analysis of the performance of the Indian stock market for six years with respect to time presents us a suitable time series arima model (1,0, i) which helps us in predicting the approximate values of the future. The analysis of the performance of the Indian stock market for six years with respect to time presents us a suitable time series arima model (1,0, i) which helps us in predicting the approximate values of the future indices. Out of the initial six different models, we choose arima(i, 0, i) as the best model based on the fact that it satisfies all the conditions for the "GOODNESS OF FIT UNLIKE THE REST"

## REFERENCES

1. Forecasting of Indian Stock( Market using Time-series APMIA Model ,Debadrita Banerjee Student, Department of Statistics St.Xavier's College, Kolkata, India(IEEE) 2014.
2. Performance Analysis of Indian Stock Market Index using Neural Network Time Series Model, D.Ashok kumar, Department of Computer Science, Government Arts College, Tiruchirappalli, India(IEEE)2013.
3. Market Prediction Model using Neural Network, S.Murugan, Department of Computer Science Alagappa Government Arts College, Karaikudi, India(IEEE)2012
4. Concentrations in Brunei Darussalam- An ARIMA Modelling Approach", Journal of Air & Waste Management and Association, Volume 54, pp 809-814.
5. Datta K.(2011)"ARIMA forecasting of Inflation in the Bangladesh Economy", The IUP journal of bank management, vol.X, No.4, pp-7-15.
6. Kumar; K Yadav; A.K.Singh, M.P; Hassan and H.Jain, V.K(2004)"Forecasting Daily Maximum Surface Ozone".
7. Dase R.K. and Pawar D.D.(2010) "Application of Artificial Neural Network for stock market predictions: A review of literature"
8. Mayankkumar B Patel, Sunil R Yalamalle (2014) "Stock Price Prediction Using Artificial Neural Network"
9. G. Preethi, B. Santhi (2012) "stock market forecasting techniques: a survey"