**Time-series Models - Forecasting Performance in the Stock Market**

Prof. Tinni Chaudhuri*, Dr. Abhijit Pandit#,  
*Assistant Professor (Grade I), Department of Statistics, Amity Institute of Applied Sciences,  
Amity University, Kolkata, Email: tinnistat1991@gmail.com  
#Assistant Professor (Grade II), Amity University, Kolkata,  
Email: abhijitpandit1978@gmail.com  

**Abstract**  
Contradicting evidence on time-series and financial analysts’ forecasting performance calls for further research in financial markets. Motivation to use time-series models rather than analysts’ forecasts stems from recent research that reports time-series predictions to be superior to analysts’ forecasts in predicting earnings for longer periods and for small firms that are hardly followed by financial analysts. The paper aims to explore performance of time series models in forecasting earnings for six firms considering historical data of 11 years from January, 2010 to December, 2020. Monthly average stock data of last 11 years for five firms namely HCL, TCS, Infosys, Reliance, Tech Mahindra and Wipro was considered from NSE site. Every company had 132 values whose graphical plotting and stationarity check was performed. Data series for each of the five companies was found to be non-stationary. After differencing each of them, the series became stationary and graphical plotting was again done. Then best suited ARIMA Model for each stationary time series was determined upon comparison of goodness of fit statistics. After choosing the best suited ARIMA model, residuals were extracted and were found to be random with no external influence whatsoever. Hence forecasting was done based on chosen model for the monthly average stock price of these top six companies of India in 2020. The paper finds that premier ARIMA family models outperform naive time-series models in terms of mean percentage errors, AIC and average ranks. The findings suggest that investors use the selected ARIMA model to form their expectations.

**Keywords**: ARIMA, time-series, forecasting, stock, financial market

**INTRODUCTION**

Financial Time Series Data (TSD) mining provides useful information for the investors, banks and insurance companies to channel their funds properly for better returns. This decision making is the primary motivation for prediction of financial TSD. Accurate multi-step ahead prediction of financial TSD becomes more difficult as the financial TSD is highly volatile. If prediction is one-step ahead preserving data trend is irrelevant. However, as the forecast horizon increases, preserving data trend becomes significant. In either case, prediction accuracy should remain high. Hence, the two basic requirements for multi-step ahead prediction model are maintaining high prediction accuracy and preserving the data trend across the prediction horizon.

Traditional models cannot meet both the requirements simultaneously. However, a hybrid model may provide scope for preserving data trend across the forecast horizon while maintaining good prediction accuracy, which motivates the research work of this paper. Most of the traditional models such as ARIMA, GARCH and ANN are applied for one-step ahead forecasting in many works of the
literature, where prediction accuracy is of major concern. However, in the present paper, we target multi-step ahead prediction, which requires preserving data trend in addition to high prediction accuracy. Such a model should account for the nature of TSD at every stage in the model.

The paper is organized in the following sections: First we present a literature survey of different prediction models existing in TSD. Next, we discuss the research methodology and the proposed ARIMA model is detailed in subsequent section. A quantitative analysis of the proposed model is discussed henceforth. The proposed and the traditional models are applied on selected NSE India data and the performance is compared. The paper ends with conclusion.

LITERATURE REVIEW

There are many kinds of research works in the area of forecasting using time series analysis. Some of the important tasks are mentioned here. A study deals with the implication of support vector machines (SVMs) regression, a novel neural network technique, in predicting the share price to examine the feasibility of SVM regression in predicting stock price. A data set related to Shanghai Stock Exchange in China has been used to test the validity of SVMs regression. The experiment depicts SVMs regression as a valuable method in forecasting the stock price (Bao et al., 2004; Pinto et al., 2020; Kumar et al., 2020). Again, a study focused on forecasting the price of Infosys Technologies, taking into consideration the previous open, close, high, and low price using different neural classifier functions like Least Mean Square, Multilayer Perceptron, Pace Regression, Linear Regression, Gaussian Processes, Simple Linear Regression, Isotonic Regression, and SMO Regression (Suresh kumar & Elango, 2011; Meher et al., 2020). Besides, a study examines the relative predictive power of ARIMA, VAR, and ECM models in predicting inflation in Nigeria. In doing this, a domestic Consumer Price Index (CPI) was lumped into the headline (all-item). Annual data from 1970 to 2010 were used.

The study examines the performance of the forecasting ability of the models. It was observed that different models performed well in different periods. While ARIMA is useful as a benchmark model, VAR for short-term forecasting and ECM are suitable for long-run forecasting (Uko & Nkoro, 2012; Bolar et al., 2017). Furthermore, in a study, the authors reviewed some of the approaches, which could be used for a stock market forecast like Hidden Markov Model, Non-linear Regression Analysis, Naive Bayes Classifier, Artificial Neural Networks, Decision Trees Classifier, Support Vector Machines, Random Forest Method, PCA (Principal Component Analysis), WB-CNN (Word Embeddings Input and Convolutional Neural Network prediction model) and CNN (Convolutional Neural Network) and finally concluded that neural network showed better results compared to other methods (Sharma & Kaushik, 2018).

RESEARCH METHODOLOGY

Any time series analysis has the following major steps in methodology in Box Jenkins approach:

a) Stationary check
b) Performing Augmented Dickey Fuller (ADF) Test / Unit Root test. If p-value of the test is less than level of significance(α), we reject null hypothesis of presence of unit root which means series is stationary.

Plotting the ACF and PACF of stationary series in order to determine the parameters of ARIMA model.

d) Fitting all possible combinations of ARIMA model by changing the values of parameter p and q in ARIMA.

e) Choosing the best ARIMA model with minimum value of AIC, since AIC is a measure of goodness of model fitting. Smaller the value of AIC, better is the fit.

f) Carrying out “Residual Analysis” of the selected model. Ljung -Box test is performed in this regard, where if p-value greater than level of significance(α) we conclude that the model fits the TSD well.
Forecasting is done based on the chosen ARIMA model for the subsequent years. There are three important parameters in the ARIMA(p,d,q) model, which are the auto-regressive order p, the difference order d and the moving average q. In the front, after one difference to determine the smooth of the sequence, we can get d=1. Next, it is necessary to determine the value of parameter p and parameter q in order to determine the final model ARIMA (p, d, q). By analyzing the auto-correlation and partial correlation graphs of time series \{y_t\}, the value of p and q can be determined.

**RESULTS AND DISCUSSION**
The original time series data of six companies (2010-2020) are shown below:

![Graphs of stock price of six companies (2010-2020)](image)

It is clear from the above graphs that all the series are non-stationary. We difference each of the series appropriate number of times till stationarity is achieved. We carry out the Augmented Dickey Fuller / Unit root test on the differenced series to be certain that they are converted to stationary series.

**Unit root test/ADF Test**
We want to test whether each of the series is stationary under this test. The null hypothesis is presence of unit root i.e. non-stationary.
As it can be seen from Table 1, the p-value corresponding to each series is less than level of significance (0.05) making accept the alternative hypothesis implying that all the differenced series are stationary.

We have also check the stationarity by ACF and PACF plots of each series.

**Model Fitting**

The cut offs (spikes) from the ACF and PACF function of the stationary series we can take various combinations of p and q to fit ARIMA (p,d,q). Here d is 1, as we difference each of the series once to make it stationary (for six companies HCL, TCS, Infosys, Reliance, Tech Mahindra and Wipro). We choose the best model looking at the value of Akaike’s Information Criteria (AIC). Lesser the value of AIC, better is the fit. The following Table 2 gives the best fitted ARIMA model for each state along with the estimated parameters and some diagnostic measures.

<table>
<thead>
<tr>
<th>COMPANY</th>
<th>TEST STATISTIC VALUE</th>
<th>P-VALUE</th>
<th>DECISION</th>
</tr>
</thead>
<tbody>
<tr>
<td>TECHMAHINDRA</td>
<td>-3.733</td>
<td>0.02447</td>
<td>STATIONARY</td>
</tr>
<tr>
<td>TCS</td>
<td>4.0532</td>
<td>0.01</td>
<td>STATIONARY</td>
</tr>
<tr>
<td>HCL</td>
<td>-3.4255</td>
<td>0.04341</td>
<td>STATIONARY</td>
</tr>
<tr>
<td>WIPRO</td>
<td>-3.3564</td>
<td>0.01</td>
<td>STATIONARY</td>
</tr>
<tr>
<td>RELIANCE</td>
<td>-7.6845</td>
<td>0.01</td>
<td>STATIONARY</td>
</tr>
<tr>
<td>INFOSYS</td>
<td>-9.058</td>
<td>0.01</td>
<td>STATIONARY</td>
</tr>
</tbody>
</table>

**TABLE 1. Unit Root Test**
### Table 2: Best fitted ARIMA model with parameter estimation and diagnostic values

<table>
<thead>
<tr>
<th>Company</th>
<th>Ljung Box test statistic</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>TECHMAHINDRA</td>
<td>37.736</td>
<td>0.899</td>
</tr>
<tr>
<td>TCS</td>
<td>28.881</td>
<td>0.9928</td>
</tr>
<tr>
<td>HCL</td>
<td>40.931</td>
<td>0.8161</td>
</tr>
<tr>
<td>WIPRO</td>
<td>46.767</td>
<td>0.6039</td>
</tr>
<tr>
<td>RELIANCE</td>
<td>48.448</td>
<td>0.5358</td>
</tr>
<tr>
<td>INFOSYS</td>
<td>28.707</td>
<td>0.9933</td>
</tr>
</tbody>
</table>

Table 3. Ljung-Box Test
We can see from Table 3, Ho is accepted in all the cases which mean that the models we have chosen for each company stock is appropriate.

2) **Residual Acf:** As we know a primary measure of goodness of fit is that errors of the fitted models should be uncorrelated. This is a desired property that ensures that the model fitted is appropriate. If the spikes of residual ACF are on both sides of the horizontal axis in a random pattern and gradually decreases to zero, we may call it an ideal situation. Hence, eventually as the lag increases the spikes get within the control band. We can see from the residual ACFs plotted in Fig. 2 that all the ARIMA models fitted are proper.

1. Techmahindra  
2. TCS  
3. Infosys  
4. Reliance  
5. HCL  
6. Wipro

**Fig. 2 Residual ACFs.**

**Forecasting**

The forecasted values of monthly average stock data of 6 companies till September 2021 is given in Table 4 for the 6 companies’ share price India along with their upper and lower control limits.
One of the most common and efficient methodologies is ARIMA modeling. The ARIMA model has been used extensively in the field of finance and economics as it is known to be robust, efficient and has a strong potential for short term prediction. ARIMA modeling of TSD. Autoregressive Integrated Moving Average (ARIMA) Model converts nearly linear time series data to stationary data before working and is one of the most popular models to predict time series data. ARIMA model has been used extensively in the field of finance and economics as it is known to be robust, efficient and has a strong potential for short-term share market prediction. In this work we have forecasted the share prices of 6 major companies of India in 2020 for the next 9 months and as we have already observed the prediction based on our chosen ARIMA model is fairly consistent. We see a significant rise in the share price of TCS in the near future, which calls for special mention. Another factor to pay attention is we do not go for a very long term prediction of economic factors are involved which play a major role in the rise and fall of market share prices. Though the companies that we have considered in this study i.e, Reliance, HCL, Tech Mahindra, Infosys, Wipro and TCS do show a steady forecast in the upcoming few years, based on their data till 2020, but investors should not discard the influence of other external factors before making their decision.

CONCLUSION
Stock and financial markets tend to be unpredictable and even illogical due to the unevenness of market volatility. Because of such characteristics financial or market data should possess a rather volatile structure which often makes it hard to find reliable patterns. Modeling such structures requires methods and algorithms capable of finding hidden pattern or trend within data and predict how they will affect them in the future. One of the most common and efficient methodology is ARIMA modeling of TSD. Autoregressive Integrated Moving Average (ARIMA) Model converts non-stationary data to stationary data before working and is one of the most popular models to predict nearly linear time series data. ARIMA model has been used extensively in the field of finance and economics as it is known to be robust, efficient and has a strong potential for short-term share market prediction. In this work we have forecasted the share prices of 6 major companies of India in 2020 for the next 9 months and as we have already observed the prediction based on our chosen ARIMA model is fairly consistent. We see a significant rise in the share price of TCS in the near future, which calls for special mention. Another factor to pay attention is we do not go for a very long term prediction of such volatile data as several socio-economic factors are involved which play a major role in the rise and fall of market share prices. Though the companies that we have considered in this study i.e, Reliance, HCL, Tech Mahindra, Infosys, Wipro and TCS do show a steady forecast in the upcoming few years, based on their data till 2020, but investors should not discard the influence of other external factors before making their decision.

References:
[31]