MODELING THE STOCK PRICE FOR AN INDIAN IT COMPANY USING TIME SERIES ANALYSIS

Arup K. Das[1], Joydeep Sen[2]

Abstract: *During the last few years, Information Technology (IT) industry has played a major role in the Indian economy. A number of large profitable Indian companies today belong to the IT sector and a great deal of investment interest is now being focused. To measure the performance, it is essential to study the behaviour of the movement of stock price. We apply the Autoregressive Integrated Moving Average (ARIMA) to model and forecast the stock price of one reputed IT company registered in Bombay Stock Exchange (BSE). For the purpose of analysis, we consider the daily data of consecutive five financial years from 2006 to 2011 of the stock-value of that IT Company.*

Diagnostic measures like R^2, Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE) indicate the adequacy of the optimum ARIMA (p, d, q) model. We conclude that the ARIMA (1, 1, 1) is adequate to explain the behaviour of the movement of stock price of that IT company and more appropriate to forecast.

Keywords: *Autoregressive Integrated Moving Average Model, Bombay Stock Exchange, Financial forecasting, Indian IT companies, Movement of stock price.*

1. INTRODUCTION

Information Technology (IT) industry has played a major role in the Indian economy during the last few years. A number of large, profitable Indian companies today belong to the IT sector and a great deal of investment interest is now focused on the IT sector.

Financial time series prediction deals with the task of modeling the underlying data generation process using past observations and using the model to extrapolate the time series into the future. Box and Jenkins [1] provided a step-by-step procedure for Autoregressive Moving Average (ARMA) or Autoregressive Integrated Moving Average (ARIMA) analysis, which is a combination of Autoregressive (AR) coefficients, multiplied by past values of the time series data and Moving Averages (MA) coefficients, multiplied by past random shocks. ARIMA model takes care of the non-stationary behavior of the stock.

Chen [2] introduced a new pre-differencing transformation for the ARIMA model for forecasting S&P 500 index volatility. Bianchi et .al [3] analyze the use of additive and multiplicative versions of Holt-winters exponentially weighted moving average models and compare it to Box-Jenkins ARIMA modeling and find that ARIMA model performs better for time series studied. In the Indian context, Mahadevan (2002) found that while forecasting 10 year government securities yield, ARIMA has a marginally better directional accuracy than that of moving average model in a static forecast. However, there is not much attempt to model individual stock price of the Indian IT companies. In this paper, we attempt to model the movement of stock price and to forecast the same using the ARIMA model.

After briefly introducing about the research background, the remaining sections of the paper is organized as follows. Section 2 explains the basics and the idea of ARIMA *(p, d, q)* model. In section 3, a brief exposition of the methodology used in the study is discussed. Section 4 presents the analysis and results along with discussion. Section 5 provides a conclusion of the study.

2. *Model:*

In statistics and econometrics, and in particular in time series analysis, an ARIMA model is a generalization of an ARMA model. These models are fitted to time series data either to better understand the data or to predict future points in the series. ARIMA model can also be applied in some cases where data show evidence of non-stationarity, where an initial differencing step (corresponding to the "integrated" part of the model) can be applied to remove the non-stationarity.

The model is generally referred to as an ARIMA(p,d,q) model where p, d, and q are integers greater than or equal to zero and refer to the order of the autoregressive, integrated, and moving average parts of the model respectively.

Assume now that the polynomial (1- $\sum_{i=1}^{p} \alpha_i L^i$) has a unitary root of multiplicity d. Then it can be rewritten as:

$$\left(1-\sum_{i=1}^{p}\alpha_i L^i\right) = \left(1+\sum_{i=1}^{p-d}\phi_i L^i\right)(1-L)^d . \qquad \ldots(1)$$

An ARIMA (p, d, q) process is given by:

$$\left(1-\sum_{i=1}^{p}\phi_i L^i\right)(1-L)^d X_t = \left(1+\sum_{i=1}^{q}\theta_i L^i\right)\varepsilon_t \qquad \ldots(2),$$

where L is the lag operator, the φ_i are the parameters of the autoregressive part, the θ_i are the parameters of the moving average part and ε_t are error terms. The error terms ε_t are generally assumed to be independent, identically distributed variables sampled from a normal distribution with zero mean and constant variance.

3. METHODOLOGY

A step-by-step stock price modeling procedure is given below.

(1) **Data Analysis:** Bases on the historical stock price data, analyze data feature, if stock price series is non-stationary series and stabilize it by lag difference.

(2) **Identification of ARIMA model and parameter estimation:** Order of integration (*d*) is chosen by successive Augmented Dickey-Fuller (ADF) test and Philips-Perron (PP) test until, the series become stationary. Orders of auto-regression (*p*) and moving average (*q*) are determined by Extended Sample Auto-correlation Factor (ESACF) method and Smallest Canonical Correlation (SCAN) method.

(3) **ARIMA model test:** Diagnostic checking for Portmanteau test is done for all the possible models (suggested by ESACF and SCAN Method) with Box-Ljung *Q* statistic. Optimum ARIMA model is chosen on the basis of having least Akaike information criterion (AIC) and Bayesian information criterion (BIC) for the models, which pass the Portmanteau test.

(4) **ARIMA model validation test:** White noise test for the ARIMA residuals is carried out.

4. ANALYSIS AND RESULT :

For the analysis of stock price we consider the daily data of consecutive five financial years from 2006 to 2011 of the stock-value of one of the reputed Indian IT company. The data is collected at a particular time interval on every working day. On Saturday, Sunday and other national holidays, Bombay Stock Exchange remains closed and also there is no transaction in stock exchange market.

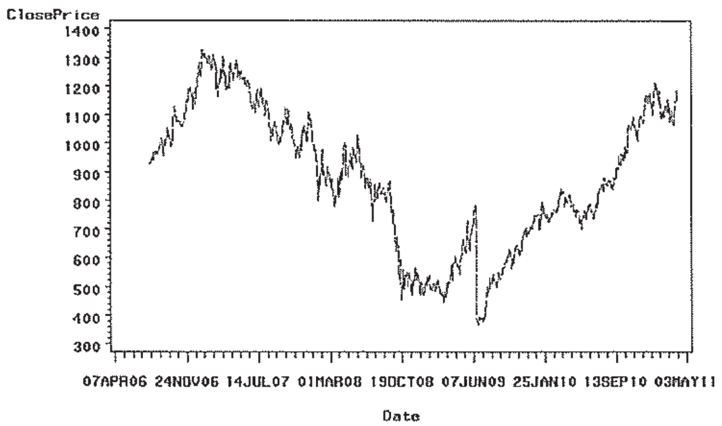

Fig. 1 Time series plot of close-price of the Indian IT company

The closing price of a security represents the most up-to-date valuation of a security until trading commences again on the next trading day. Closing price provides a useful marker for investors to assess changes in stock prices over time. The closing price of one day can be compared to the previous closing price in order to measure market sentiment for a given security over a trading day. So we consider only closing price of the stock for our analysis.

The value of ADF test statistic for testing stationarity = -1.35 > -1.645 = tabulated τ at level 0.05, indicates the non-stationarity of the original data. To achieve mean-stationarity, we differentiate the series for lag 1. For the first differenced series, ADF test statistic = -24.78 < -1.645 = tabulated τ at level 0.05, PP statistic = -33.97 < -1.645 = tabulated τ at level 0.05 and p-value < 0.0001 indicate that the first differenced series becomes mean stationary. Now we check whether at d = 1, the series is over differentiated or not. Applying Variate Difference method we see that, d=1 is correct. We determine the correct order of p and q by ESACF, SCAN method.

Table 1. ARMA (p+d, q) Tentative Order Selection Tests

ESACF Method			SCAN Method		
$p + d$	q	BIC	$p + d$	q	BIC
1	1	6.2506	1	1	6.2504
2	1	6.2556	2	2	6.2611

(5% Significance Level)

From the table 1, we select the following three ARIMA models and compare them.

a) ARIMA (0, 1, 1)

b) ARIMA (1, 1, 1)

c) ARIMA (1, 1, 2)

Along with estimating the parameters by Maximum Likelihood (ML) method, we carry out the diagnostic test for these three ARIMA models. We carry on the Portmanteau test for ARIMA residuals with Box-Ljung Q Statistics for testing the randomness of the ARIMA residuals.

Table 2. Portmanteau test for the ARIMA models

Model	**Value of Q statistic at lag 6 with degrees of freedom (d.f.)**	**Tabula--ted chi-square value at level 0.05**	**Comment about the acceptance or rejectior the H_0: randomness of the ARIMA residu**
ARIMA(0,1,1)	8.16 with d.f. 5	11.07	H_0 is accepted
ARIMA(1,1,1)	3.12 with d.f. 4	9.49	H_0 is accepted
ARIMA(1,1,2)	8.07 with d.f. 3	7.81	H_0 is rejected

Table 3. AIC and BIC values of the ARIMA models

Model	AIC	SBC
ARIMA(0,1,1)	10519.62	10533.64
ARIMA(1,1,1)	10514.11	10531.08

Based on analysis from table 2 and 3, we select ARIMA (1,1,1). Table 4 depicts the ML estimate of the parameters.

Table 4. Maximum Likelihood estimate of ARIMA (1, 1, 1)

Parameter	Estimate	Standard Error	t Value	Approx Pr > |t|	Lag
MA1,1	0.88442	0.08960	9.87	<.0001	1
AR1,1	0.84724	0.10241	8.27	<.0001	1
Variance Estimate			512.82		
Std Error Estimate			22.65		
AIC			10514.11		
SBC			10524.22		

The graph of the actual vs predicted close-price of the IT company is given figure 2. For ARIMA (1,1,1), Fisher's Kappa = 5.1107, Bartlett's Kolmogorov-Smirnov Statistic = 0.0302 and its approximate p-value = 0.6663 indicate white noise property of the residuals. As we get, R-square = 0.9913, Standard Error of Estimate = 22.65, Mean Absolute Error (MAE) = 14.67 and Mean Absolute Percentage Error (MAPE) =1.83%, we can say that the model is adequate. The kurtosis of the ARIMA residuals = 3.05 indicates asymptotic normality of the residuals.

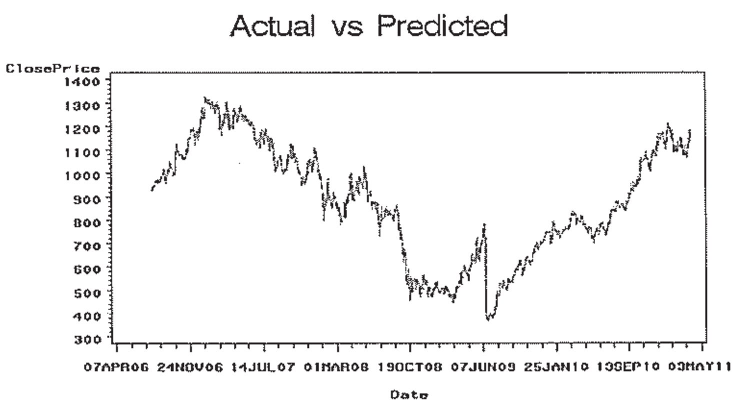

Fig. 2 Actual vs Predicted close-price of that IT company

5. CONCLUSION

Improving forecasting, especially time series forecasting accuracy is an important yet often difficult task for the decision makers in financial areas. Despite the numerous time series models available, the research for improving the effectiveness of financial forecasting models has never stopped.

Our proposed ARIMA model has the unique capability in linear modelling in order to identify magnify the existing linear structure of the data and can takes care of the non-stationarity problem. We conclude that the ARIMA(1,1,1) is adequate to explain the behaviour of the movement of daily stock price of that company and appropriate for forecasting.

6. ACKNOWLEDGMENT

Authors are grateful to the Bombay Stock Exchange for their assistance in obtaining the data.

7. REFERENCES

1. P. Box and G.M.Jenkins, Time Series Analysis: Forecasting and Control. Holden-day Inc., San Francisco, CA , 1976.
2. Chen, "Forecasting the S&P 500 index volatility", Int. Rev. Eco. Finance., vol. 6(4), 1997, pp 391-404
3. Bianchi, J. Jarret and R. C Hanumara, "Improving forecasting for telemarketing centers by ARIMA modeling with intervention", Int. J. Forecasting., vol.14(4), 1998, pp. 497-504.
4. Peter J. Brockwell and Richard A .Davis, Introduction to Time Series and Forecasting, Second Edition, Springer Text in Statistics. New York, 2002.
5. J.M Bates and C.W.J Granger, "The combination of forecasts", Oper. Res. Q., vol. 20 , 1969 pp. 451-468.
6. C. Chatfield , " What is the 'best' method of forecasting?", J. Appl. Statist., vol. 15, 1988, pp .19-39.
7. C.Chatfiled, "Model uncertainty and forecast accuracy", J. Forecasting., 15, 1996, pp. 495-508

* * * * *

Author 1
Address : Kolkata
Affiliation : SQC & OR Division, Indian Statistical Institute, Kolkata
akdas@isical.ac.in

[2]*Author 2*
Address : Kolkata
Affiliation: Indian Statistical Institute, Kolkata
joydp.sen@gmail.com