Journal of Applied Econometrics and Statistics
Vol. 2, No. 1-2, 2024, pp. 97-115
https://DOI:10.47509/JAES.2024.v03i01-2.06

ARF INDIA
Academic Open Access Publishing
www.arfjournals.com

Forecasting Future Stock Price Volatility: Error Variance Estimation

T. Lakshmanasamy

Formerly ICSSR Senior Fellow and Professor, Department of Econometrics, University of Madras, Chennai. E-mail: tlsamy@yahoo.co.in

Received: 18 July 2024; Revised: 19 August 2024; Accepted 28 August 2024; Publication: 30 December 2024

Abstract: In any stock market, the stock prices are generally volatile over time. While stock prices either increase or decrease gradually in short periods, the fluctuations are wide and more persistent over long periods. This paper analyses the effects of such short-period and long-period volatility on stock prices. Using error variance *i.e.* volatility in residuals, specifically volatility clustering, the future volatility in stock prices is forecasted. Using data on the stock prices of TATA Steel Limited, a listed company in the NSE, for the period January 1, 2021- December 31, 2023, the effects of short-period and long-period stock price fluctuations on daily stock prices and volatility are predicted for the next 69 days, from January 1 to April 13, 2024. The stock prices are predicted first by the ARIMA model, and then the future stock price volatility is predicted by applying the GARCH and EGARCH models on the resultant residuals. The EGARCH fitting shows that the long-period fluctuations have a significant effect on the future stock price volatility relative to the GARCH fitting. The comparison of EGARCH forecasts with the actual stock price fluctuations from January 1-April 13, 2024, shows that the long-period stock price volatility is more reliant than the short-period volatility in forecasting future stock price volatility as well as the stock prices.

Keywords: Stock price volatility, heteroscedasticity, error variance, volatility clustering, asymmetry, leverage, ARIMA, GARCH, EGARCH, forecasting.

INTRODUCTION

The prices of shares in all stock markets fluctuate both during the time of transactions and over time. Fluctuations in stock prices include both positive and negative changes. The stock price of any product is affected by a host of factors such as inflation, demand for that good in the market, government policies, changes in budget, the worth of the organisation or company that manufactures that particular good, etc. Among all the factors that cause stock price fluctuations, the performance of the company whose shares are traded matters a lot for the stock price as well as its volatility. Further, the stock prices and volatility are related to systematic risk as well as unsystematic risk in the stock market. The investor's timing of buying and selling of shares is also influenced by calendar effects like the January effect, Friday the 13th effect, and the first half of the month effect. The volatility of the stock price is generally measured by the variability of the stock prices over time, and the common measure of volatility is the standard deviation of returns on the stocks.

The Indian stock market covers major sectors of the Indian economy, including financial services, information technology, automobiles, energy, metal, engineering, etc., and offers investment managers a vibrant exposure to the Indian market. In India, there are two major stock exchanges: the Bombay Stock Exchange (BSE) operating from Mumbai and the National Stock Exchange of India (NSE) operating from New Delhi. The performance of stocks in stock exchanges is measured by some indices. The two such indices in India are the BSE30, commonly the SENSEX, of the BSE and NIFTY50 of the NSE. The NIFTY50 stock index is widely considered the benchmark and barometer for the capital markets in India.

In trading stocks in a day, the stock prices are bound to fluctuate between the open price and the close price of the day due to various reasons. It is also possible that the price of a stock experienced in a day is influenced by its previous price, even a long past ago. If such is the case then today's price may have an effect on the price that comes into effect tomorrow or later. Thus, there may be lagged effects of own price in the current stock price. It is also possible that current stock price volatility is a reflection of past fluctuations. Therefore, past stock prices and its volatility may be crucial to forecasting future prices and their volatilities. Econometrically, based on previous prices, the future price of a stock can be forecasted and also its volatility can be forecasted. Among the many econometric models of forecasting, autoregressive integrated moving average (ARIMA), autoregressive conditional heteroscedasticity (ARCH) and generalised autoregressive conditional heteroscedasticity (GARCH) models are commonly used to forecast future stock prices and their volatilities. While the ARCH and ARIMA models assume constant common variance, the GARCH model takes into account the time-varying conditional variance of stock prices. The conditional variance includes past variances in the autoregressive term and the moving average term is the square of residual from the autoregression of present variance on the past variance.

This paper analyses the impact of long-term and short-term stock price volatility on the future stock price as well as on the variance of the forecasted stock price. Specifically, this paper forecasts the future stock price using datasets for long and short time periods and forecasts the variance from the long-term and short-term stock price fluctuations. The empirical analysis is based on the daily stock prices of TATA Steel Limited, a listed company in the NSE that manufactures metal steels. Tata Steel Limited, a Tata Group subsidiary, is the second largest steel manufacturing company in India, after Steel Authority of India Limited, a public sector undertaking in India, with a global presence. The daily data on the stock prices of TATA Steel for the period January 1, 2021 to April 13, 2024, from the National Stock Exchange of India has been used. Empirically, this paper follows the Box-Jenkins methodology for forecasting. The residuals of the ARIMA model are used in the GARCH and EGARCH models for an understanding of the error variance effect on the forecast of future volatility in stock prices.

LITERATURE REVIEW

In many analyses of stock market performance and stock prices, the ARIMA model based on the Box-Jenkins method is used for forecasting. The ARIMA method easily handles the nonstationary data, an important nature of stock price series. However, the ARIMA model forecasts only the future stock prices, but not the variance fluctuations. The ARIMA-GARCH method forecasts the future values of stocks as well as the fluctuations in the variance of stock prices.

Maity and Chatterjee (2012) apply the ARIMA of order (1,2,2) for the period 1959 to 2011 for forecasting the GDP of India for the next ten years, 2012 to 2021. The forecasted GDP shows an increasing trend and its rate of growth rates shows a decreasing trend. In the maximum likelihood estimates, the coefficient of AR terms is negative and less than 1 and the coefficients of MA are more than 1. The statistical validity of the model is checked by modified Ljung-Box statistics. The estimates show a parsimonious model with only one AR coefficient and one MA coefficient with statistical significance. Therefore, they argue that the ARIMA model is very effective not only in forecasting GDP but also in predicting the growth rate of GDP in India.

Guha and Bandyopadhyay (2016) apply the ARIMA model to the November 2003 to January 2014 nonstationary gold price data for forecasting the gold price. They compare various statistics of fitness like mean absolute error, mean absolute percentage error and root mean squared error of ARIMA models of different orders. They conclude that the ARIMA (1, 1, 1) forecast is the most accurate forecast for the data. The forecasted gold prices show an increasing trend.

Ashik and Kannan (2017) apply the ARIMA model to the 2015 NSE Nifty 50 closing price to forecast future stock prices. Among all the ARIMA models, the ARIMA (0,1,1) forecast is the most precise forecast, with the lowest BIC value and small mean absolute percentage error. The closing stock price of Nifty 50 shows a trend with slow decreasing fluctuations for future trading days.

The assumption of common variance may not be satisfied commonly in all the time series errors. Fluctuations in errors may produce volatility in variance also. In many instances, the variance may be conditional and vary over time. Under such conditions, the ARIMA modelling is inappropriate. To overcome the common variance assumption, Engle (1982) proposes a time-varying conditional variance model, the ARCH model. A generalisation of the ARCH model, the GARCH model is proposed by Bollerslev (1986). In the GARCH model, the periods of fluctuations are clustered and the volatility of future stock prices is predicted. The GARCH model predicts future variances on the basis of AR, MA, ARMA or ARIMA forecasts of the variance of residuals.

There may also be an asymmetry in volatility due to large positive and negative stock returns. The positive stock returns could arise because of good news when there is calmness in the financial market and the negative returns on stocks may arise because of bad news in a period of volatile financial market. When there is an asymmetry in volatility and leverage effect, the GJR-GARCH or TGARCH is used for forecasting the variance (Glosten, Jagannathan and Runkle, 1993).

Ahmad *et al.* (2015) use a hybrid of linear ARIMA and GJR-GARCH to model and forecast the Malaysian gold price. They compare the TARCH forecasts with ARIMA forecasts for forecasting accuracy. The gold prices are forecasted using the best fit ARIMA model whose order is (2,1,2). Then, the residuals of the forecasted values are subjected to TGARCH analysis. The ARCH-LM test is applied to the residuals for ARCH effects. Then, the GJR-GARCH model of order (1,1) has been applied to forecast the gold price. Based on the lowest AIC values, the ARIMA (2,1,2)-GJR-GARCH (1,1) hybrid model is shown to perform better than the ARIMA model in forecasting gold prices. Further, in terms of forecasting, the ARIMA-GJR-GARCH produces lower in-sample and out-sample mean absolute percentage errors (MAPE) compared to those of the ARIMA model.

Yaziz *et al.* (2016) use the 5-day-per-week frequency data on the daily gold price for 40 days from November 26, 2005 to January 18, 2006to forecast the gold price in Malaysia applying the ARIMA-GARCH hybrid model. The ARIMA (1,1,1) has been fitted to the 35 observations, and the remaining 5 observations are used to check the accuracies of the forecasts. To the residuals of these forecasts, GARCH has been fitted to forecast the variance of the prices. The results show that the ARIMA (1,1,1)-GARCH (0,2), with low mean square error and mean absolute error, is the most efficient model which produces optimum results.

Epaphra (2017) apply the GARCH and EGARCH models to the exchange rate of the Tanzanian Shilling and UD$, in order to study the volatility in exchange rate in Tanzanian. The data used is from January 4, 2009 to July 27, 2015. The variance is modelled using GARCH (1,1) and the asymmetry and leverage effects are captured by EGARCH (1,1) models. The negative coefficient of asymmetric volatility signifies less volatility with respect to positive shocks relative to negative shocks. The GARCH (1,1) model is a good fit as its root mean square error is low.

DATA AND METHODOLOGY

This paper uses the daily data of the listed TATA Steel Limited stock prices from the National Stock Exchange of India. The time period considered is from January 1, 2021 to April 3, 2024. The daily data on stock prices from January 1, 2021 to December 31, 2021 is used for short-run variance forecasting, and data from January 1, 2021 to December 31, 2023 is used for long-run variance forecasting. The forecast accuracies are validated with the data from January 1, 2024 to April 13, 2024, the next 69 days. The daily stock price data of TATA Steel Limited is collected from the NSE website that contains information on open, close, high, low, previous close, last stock prices, volume weighted average price (VWAP), number of trades, total traded quantity, total deliverable quantity, percentage of deliverable quantity to traded quantity, turnover, etc.

EMPIRICAL METHODOLOGY

The VWAP *i.e.* volume weighted average price has been taken for fitting the ARIMA model, since unlike open, close and last prices, VWAP takes into account all the prices that existed throughout the day and the total trades that have taken place for different prices.

$$\text{VWAP} = \frac{\Sigma\text{Stock price} \times \text{no. of stocks bought at that price that day}}{\text{Total no. of stocks traded for that day}}$$

The future stock prices are forecasted by the ARIMA model and the GARCH is fitted on the residuals of the forecasts. First, to check for stationarity, the data are plotted at levels and the Augmented Dickey-Fuller (ADF) test is performed on the levels data. The time series is then differenced and the stationarity of the differenced series is checked by applying the ADF test yet again. After differencing, the series achieves stationarity. Based on the Autocorrelation Function (ACF) and the Partial Autocorrelation Function (PACF) plots of the differenced series, tentative orders of the autoregressive and moving average terms are taken just to get a point to begin from. Then, based on the significance of coefficients and lowest AIC values, the best-fit model is identified. The variables in the resultant estimating equation are in differenced forms. Hence, the difference is eliminated from the equation by using basic sum-difference arithmetic and then forecasting is carried out. The Breusch-Godfrey LM and ARCH LM tests are applied to the residuals of the forecasts to test for serial correlation and ARCH effects respectively. Once the existence of ARCH and GARCH effects is confirmed, the GARCH model is applied, thereby obtaining forecasts of the varying conditional variance *i.e.* volatility. After checking for the significance of coefficients, the EGARCH model is applied to the residuals of the forecasts generated from the data set taken for a long time period.

BOX-JENKINS METHODOLOGY

A time series data ordinarily does not reveal itself what process it follows– AR or MA or ARMA or ARIMA. Even if the appropriate process is known, the orders of the model, the p, r, q or d of the process are not easily identifiable. Box-Jenkins (1976) addresses these issues. The B-J methodology proceeds in four steps:

(i) identification,

(ii) estimation,

(iii) diagnostic checking, and

(iv)forecasting.

For, example, the ARIMA(p, d, q) could be fitted to the time series if only the orders of the autoregressive process (p), integration (d), and moving average process (q) are identified.

First, the unit root test has to be applied on the time series before and after differencing to identify the order of integration, d. If the ADF test on the undifferenced series reveals no unit root *i.e.* the series is stationary at levels, then the order of integration is zero. If the ADF test shows that the undifferenced series has a unit root and the differenced series has no unit root *i.e.* the series is stationary at the difference, the order of integration is one. The differencing process continues until the series achieves stationarity *i.e.* till the unit root is eliminated from the time series. Thus, the order of integration, d, is identified as that level of differencing at which the series is stationarised by the elimination of the unit root.

The ARIMA is the fitted on the difference-stationarised time series for determining the AR (p) or MA (q) in order to correct for any autocorrelation present in the differenced series. The orders of AR and MA terms of the series are identified by the autocorrelation function (ACF) and partial autocorrelation (PACF) plots of the differenced series. These plots show the correlation of the series with its own lags. In time series data, the partial correlation propagates to higher-order lags, and hence the correlation between the lags also propagates. When the series is not fully differenced *i.e.* if the PACF of the differenced series shows a cutoff and/or the lag-1 autocorrelation is positive, an AR (p) term may be added to the model indicating the number of AR terms. On the other hand, if the series is over differenced *i.e.* the ACF of the differenced series indicates a cutoff and/or the lag-1 autocorrelation is negative, an MA (q) term indicating the lag at the ACF cut-off may be added to the model. Once the model order (p, d, q) is identified, the ARIMA model is fitted on the integrated series of order d, autoregressive terms p, and moving average terms q, applying the ordinary least squares regression.

In the ARIMA model of stock prices, the current price is expressed in terms of a sum of past prices Y_{t-p}'s and the sum of moving average terms or past error terms u_{t-q}'s. Each u_{t-q} is obtained by regressing Y_{t-p} on Y_{t-p-1} ($p = 0$, 1, 2…n). As the correlogram and partial correlogram are the basis for determining the order of the AR and MA terms, there may also exist many other models that can fit better to the data. The best fit ARIMA model among different orders of the autoregressive and moving average terms is generally chosen on the basis of the significance of coefficients and on certain criteria such as log-likelihood, Bayesian Information Criteria (BIC), Akaike Information Criteria (AIC) or Schwarz-Bayesian Information Criteria (SBIC).

After choosing the best fit ARIMA model, the next step is to check if there is more information available *i.e.* to check if there are any more significant autocorrelations and partial autocorrelations at any lags present. In essence, the diagnostics is to confirm white noise residuals. In correlogram and partial correlogram analysis, the Box-Pierceq-statistic or Ljung-Box q-statistic tests the joint hypothesis that all the autocorrelations up to certain lags are simultaneously equal to zero.

The best fit ARIMA model obtained through the foregoing identification, estimation and diagnostic checking steps can now be used for forecasting future stock prices. The estimating equation consists of Y_{t-p} autoregressive and u_{t-q} moving average terms based on the order. Thus, the ARIMA model is a self-determining model as there are only current and past values of the data series that depend on the own autoregression of the variable and the moving average of the errors, and no exogenous variables are there in the model.

ARIMA MODEL

Generally, the autoregressive model AR(p) is specified as:

$$Y_t = a + \gamma_1 Y_{t-1} + \ldots + \gamma_p Y_{t-p} + u_t \quad (1)$$

where Y_t is a finite linear sum of its past values, u_t is the random shock or white noise term identically and independently distributed, $u_t \sim$ NIID (0, σ^2), γ_i ($i = 1, ..., p$) are the parameters of the model, and Y_t is stationary.

The moving average model $MA(q)$ is specified as:

$$y_t = u_t - \gamma_1 u_{t-1} - \lambda_2 u_{t-2} + u_{t-q} \tag{2}$$

where Y_t is a linear weighted sum of the current and past values of the random shock series, and the l_j ($j = 1, ..., q$) are the moving average parameters. When the series is nonstationary at level, the series Y_t is reduced to stationarity by differencing:

$$\Delta Y_t = Y_t - Y_{t-1} \tag{3}$$

Then, the AR (1) model for the differenced series is specified as:

$$\Delta Y_t - \gamma_1 \Delta Y_{t-1} - ... - \gamma_p \Delta Y_{t-p} = u_t \tag{4}$$

The combination of AR and MA models along with an appropriate degree of differencing (integration), the ARIMA (p, d, q) model is specified as:

$$Y_t = a + \gamma_1 Y_{t-1} + \gamma_2 Y_{t-2} + ... + \gamma_p Y_{t-p} + u_t - \lambda_1 u_{t-1} - \lambda_2 u_{t-2} - ... - \lambda_q u_{t-q} \tag{5}$$

or

$$\Delta Y_t - \gamma_1 \Delta Y_{t-1} - ... - \gamma_p \Delta Y_{t-p} = u_t - \lambda_1 u_{t-1} - \lambda_2 u_{t-2} - ... - \lambda_q u_{t-q} \tag{6}$$

Using the backward shift operator B, a compact ARIMA model is specified as:

$$Y_{t-1} - Y_{t-2} = BY_t - B^2 Y_t \tag{7}$$

$$\Rightarrow \quad Y_{t-1} = B(1 - B)Y_t;\ \Delta Y_{t-2} = B^2(1 - B)Y_t;\ (1 - B)Y_t = Y_t - Y_{t-1} \tag{8}$$

where $B^d Y_t = Y_{t-d}$.

The general ARIMA (p, d, q) model, with d degree of differencing, is expressed as:

$$(1 - \gamma_1 B - \gamma_2 B^2 - \gamma_3 B^3 - ... - \gamma_p B^p)(1 - B)Y_t = (1 - \lambda_1 B - \lambda_2 B^2 - ... - \lambda_q B^q)u_t \tag{9}$$

The ARIMA model can be estimated by the OLS method by regressing the differenced series on the autoregressive and moving average terms. However, when the data is fitted with the model, errors are bound to arise showing the deviation of the fitted values from the actuals. The estimated residuals are calculated as:

$$\hat{\mu}_t = Y_t - \hat{y}_t \tag{10}$$

Then, the GARCH model is applied to estimate the error variance *i.e.* volatility in residuals. The common variance of errors is tested by applying the ARCH-LM test.

GARCH MODEL

When the errors of the time series exhibit time-varying errors, then the OLS method cannot be applied. Engle's (1982) ARCH modelling is applied when the time series exhibits time-varying conditional variance. Bollerslev's (1986) GARCH modelling is applied when the series exhibits volatility clustering to predict future volatility. The GARCH model incorporates the variance and variance forecast of the previous periods in the forecast of future variance. In the standard GARCH model, the past volatility and variance are symmetric.

The basic structure of the symmetric and normal GARCH (p, q) model is specified as (Brooks, 2008):

$$Y_t = \gamma_0 + \gamma_1 Y_{t-1} + \ldots + \gamma_q Y_{t-q} + u_t \tag{11}$$

where the error term u_t, conditional on information of period t–1, is distributed as:

$$u_t \sim [N(0, (\gamma_0 + \gamma_1 u_{t-1}^2 + \lambda\sigma_{t-1}^2)] \tag{12}$$

The error variance follows the ARCH (1) process. The error variance depends on the squared error as well as its conditional variance in the previous period. As the error variance is not directly observed, Engle (1982) suggests the conditional variance for GARCH modelling:

$$u_t = v_t \sigma_t^2 \tag{13}$$

$$\sigma_t^2 = \delta + (\gamma_1 u_{t-1}^2 + \gamma_2 u_{t-2}^2 + \ldots) + (\lambda_1 \sigma_{t-1}^2 + \lambda_2 \sigma_{t-2}^2 + \ldots) \tag{14}$$

$$\sigma_t^2 = \delta + \Sigma\gamma_i u_{t-q}^2 + \Sigma\lambda_j \sigma_{t-p}^2 \tag{15}$$

where $v_t \sim$ N (0, 1) and $\delta = \gamma_0 (1 - \lambda_1)$. The GARCH term is the conditional variance s^2, where order p represents the forecast variance of the last period. The ARCH term u^2 represents the previous period volatility, the q lags of the squared residual from the mean equation.

EGARCH MODEL

The disadvantage of the GARCH model is that it assumes the parameters to be non-negative and symmetry in residuals. Nelson and Cao (1992) propose an Exponential Conditional Heteroscedasticity (EGARCH) model that takes into account the leverage effect *i.e.* asymmetry in the residuals induced by big positive and negative changes. The EGARCH model is specified as:

$$\log \sigma_t^2 = \omega + \Sigma_{k=1}^{q} \alpha_k g(z_{t-k}) + \Sigma_{k=1}^{p} \beta_k \log \sigma_{t-k}^2 \tag{16}$$

$$g(z_t) = \alpha z_t + \beta[\,|z_t| - E|z_t|\,] \tag{17}$$

$$Zt = \frac{u_t}{\sqrt{\alpha_t^2}} \tag{18}$$

If $z_t \sim N(0, 1)$, then $E|z_t| = \sqrt{\frac{2}{\pi}}$ (19)

Inserting equations (18) and (19) into equation (17) gives:

$$g(z_t) = \alpha \frac{s_t}{\sqrt{\alpha_t^2}} + \beta\left[\frac{|u_t|}{\sqrt{\alpha_t^2}} - \sqrt{\frac{2}{\pi}}\right] \quad (20)$$

Substituting one period-lagged form of equation (20) in equation (16) yields:

$$\log \sigma_t^2 = \omega + \lambda \ln \sigma_{t-1}^2 + \theta \frac{u_{t-1}}{\sqrt{\sigma_{t-1}^2}} + \gamma\left[\frac{|u_{t-1}|}{\sqrt{\alpha_{t-1}^2}} - \sqrt{\frac{2}{\pi}}\right] \quad (21)$$

where $\sigma_t^2 = \gamma_0 + \gamma_1 u_{t-1}^2 + \lambda_1 \sigma_{t-1}^2 + \theta u_{t-1}^2 I_{t-1}$ and I being the information asymmetry.

Thus, the EGARCH model has the advantage of less restrictiveness, since σ_t^2 as modelled is always positive even with negative parameter values. The persistence of conditional volatility is measured by the term λ and the asymmetry or leverage effect is measured by q. For large λ, the volatility takes a long time to decay. When $\theta = 0$, the model is symmetric, $\theta < 0$ implies less volatility with positive shocks than with negative shocks, and $\theta > 0$ implies more volatility due to good news than bad news. The γ measures the symmetric effect or the GARCH effect of the model. Equation (21) is an EGARCH model of the first order, where the conditional variance σ_t^2 is asymmetric with respect to u_{t-k}, the lagged disturbances. The EGARCH model uses the logarithmic value of the conditional variance.

EMPIRICAL ANALYSIS

The stock price of TATA Steel Limited in NSE has experienced fluctuations over the data period as revealed by the VWAP data at levels showing no mean reversal phenomenon. The stationarity of the series is checked by plotting VWAP data at first difference which exhibits the phenomenon of mean reversal *i.e.* the graphs are frequently cutting the mean line. The data has become stationary after differencing it once.

Towards the fitting of the ARIMA model for the data, identification of the orders (p,d,q) of autoregressive, integration or differencing and moving average terms are to be determined by performing the standard tests on the data series. The Augmented Dickey-Fuller test for stationarity is performed both at the levels and on the first difference. The ADF test is performed on the data by estimating the regression of the form:

$$\Delta Y_t = \beta_0 + \beta_t + \delta Y_{t-1} + \Sigma\beta_i \Delta Y_{t-k} + u_t \qquad k = 1, \ldots, T \quad (22)$$

where u_t is a pure white noise error term and Δ is the first difference. The Akaike Information Criteria (AIC) is used to determine the number of lagged difference terms to be included in the model to avoid serial correlation in the error term. In ADF, the test is whether $\delta = 0$ *i.e.* if $\rho = 1$ for the presence of unit root in the time series. Table 1 presents the ADF test results showing that the series is nonstationarity at levels. Hence, the series at levels has a unit root, which is also revealed by the time series plot of the differenced series. At first difference, the series has become stationary, as the t-statistics > 0.05 *p*-value. Thus, the data series is integrated of order 1 *i.e.* *I* (1).

Table 1: ADF Stationarity Test

At level		*At first difference*	
t-statistic	*Probability*	*t-statistic*	*Probability*
–2.053651	0.2639	–35.21691	0.0000

The orders of AR and MA terms are determined by checking the ACF and PACF of the data series. The ACF and PACF correlograms of the differenced short-period and long-period data series, as revealed in Figures 1 and 2, show a significant autocorrelation at the first lag in both ACF and PACF.

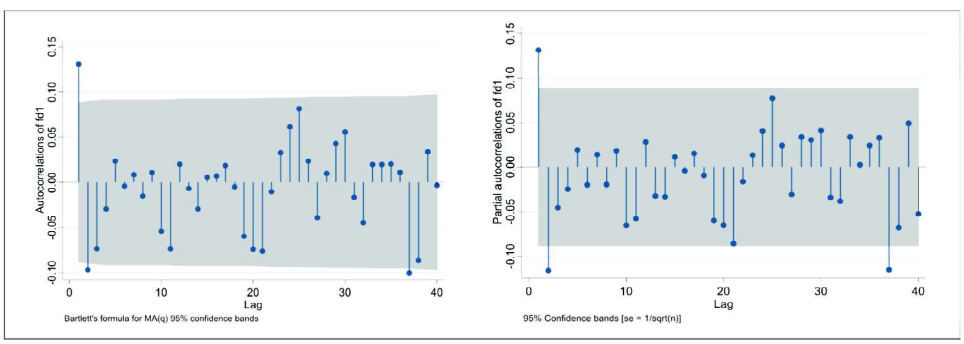


Figure 1: Short-period Autocorrelations of First Differenced Series

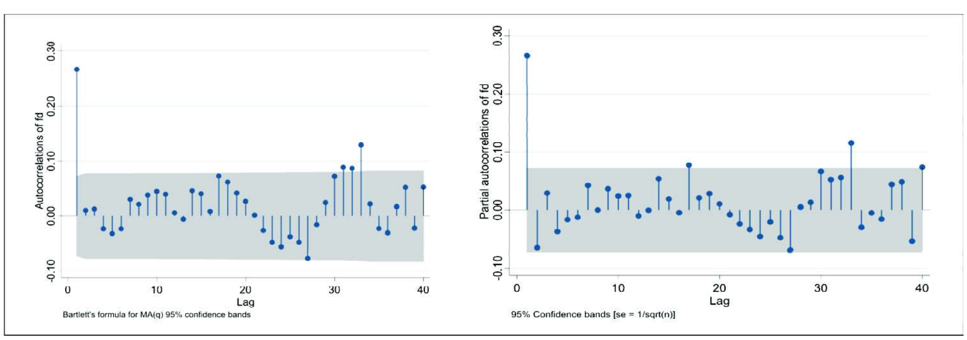


Figure 2: Long-period Autocorrelations of First Differenced Series

Therefore, based on the correlogram and unit root test, the (p, d, q) order of the model has to be (1, 1, 1), due to the presence of autocorrelation at the first lag itself. However, the correlogram can indicate only the orders and does not identify the appropriate order for the ARIMA model. Hence, the AIC has to be applied for the identification of the most appropriate order to be chosen for model fitting. The best-fit model is the one with the lowest AIC value. Table 2 presents the AIC and SIC values for various orders of both short-period and long-period time series. Based on the significance and the lowest AIC, the orders (1, 1, 2) and (3, 1, 3) are identified as the best-fit models for short-run and long-run data series respectively.

Table 2: AIC for Orders (p, d, q)

Order (p, d, q)	*Short-period*		*Long-period*	
	AIC	*SIC*	*AIC*	*SIC*
(1,1,1)	6.9628	6.9883	6.7672	6.7796
(1,1,2)	6.9571	6.9911	6.7684	6.7849
(1,1,3)	6.9631	7.0055	6.7699	6.7907
(2,1,1)	6.9587	6.9927	6.7683	6.7848
(2,1,2)	6.9627	7.0051	6.7684	6.7891
(2,1,3)	6.9667	7.0176	6.7700	6.7949
(3,1,1)	–	–	6.7685	6.7892
(3,1,2)	–	–	6.7698	6.7946
(3,1,3)	–	–	6.7629	6.7920

The OLS regression estimates of the best fit ARIMA model for the first differenced series on the autoregressive and moving average terms are presented in Table 3.

Table 3: OLS Estimates of the Best Fit ARIMA Model

Data period	*Parameter*	*Coefficient*	*Standard error*	*t-statistics*	*p-value*
Short run	AR(1)	0.877*	0.0721	12.1642	0.000
	MA(1)	–0.733*	0.0814	–9.0044	0.000
	MA(2)	–0.187*	0.0451	–4.1503	0.000
	Constant	1.016*	0.2362	4.3023	0.000
Long run	AR(1)	–1.715*	0.1127	–15.222	0.000
	AR(2)	–1.302*	0.1658	–7.8511	0.000
	AR(3)	–0.220**	0.1093	–2.0143	0.044
	MA(1)	1.948*	0.1037	18.7918	0.000
	MA(2)	1.668*	0.1539	10.835	0.000
	MA(3)	0.454*	0.1025	4.4261	0.000
	Constant	0.247	0.2418	1.0203	0.308

Note: *significant at 1, 5% levels

Thus, the short-period model is:

$$\Delta y_t = 1.016 + 0.877\Delta y_{t-1} - 0.733u_{t-1} - 0.187u_{t-2} \tag{23}$$

The long-period model is:

$$\Delta y_t = 0.247 - 1.715\Delta y_{t-1} - 1.302\Delta y_{t-2} - 0.220\Delta y_{t-3} + 1.948u_{t-1} + 1.668u_{t-2} + 0.454u_{t-3} \tag{24}$$

Since the regression is on the differenced series, the resultant forecasted prices would be in differenced forms as well. Hence, using ordinary sum-difference arithmetic, the differenced terms from the equations are eliminated to get estimated equations in terms of actual values of prices. Then, the stock prices are forecasted in the base form. Thus, from equation (23), the forecasted stock price at time *t* is:

$$y_t = 1.016 + 0.123y_{t-1} + 0.877y_{t-2} - 0.733u_{t-1} - 0.187u_{t-2} \tag{25}$$

From equation (24), the forecasted stock price at time *t* is:

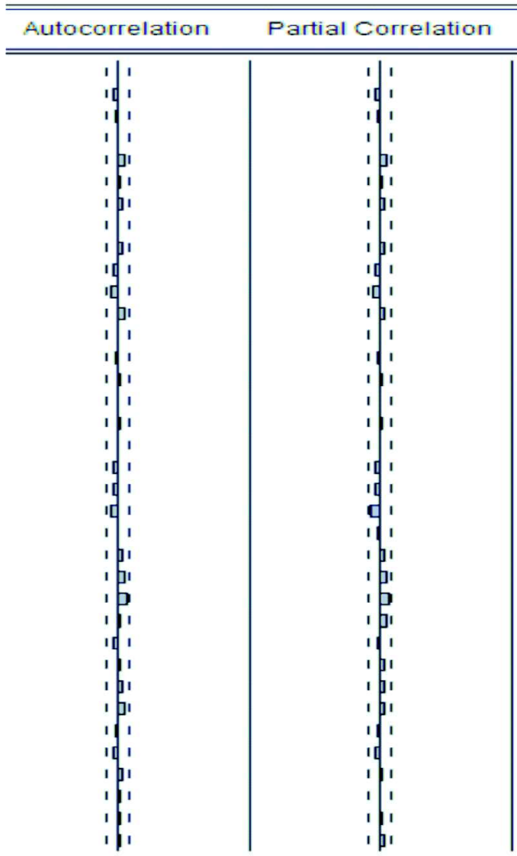
$$y_t = 0.247 - 715y_{t-1} - 0.414y_{t-2} - 1.082y_{t-3} + 1.948u_{t-1} + 1.668u_{t-2} + 0.454u_{t-3} \tag{26}$$

Therefore, based on short data and long data, equations (25) and (26) respectively are used in the forecasting of the TATA Steel stock prices.

Before forecasting the stock prices, following the Box-Jenkins methodology, the residuals of the best-fit models are tested if they are white noise terms. Figures 3 and 4 show the correlograms of residuals generated by the short-period and long-period models respectively. As can be seen from the correlograms, no significant spikes can be noticed in both ACF and PACF, implying that the residual of the identified ARIMA models is white noise, and no further information is available. Hence, there is no need to consider AR (*p*) and MA (*q*) any further.

Using the identified models for forecasting, ARIMA models have been fitted for the long and short periods, and TATA Steel stock prices have been predicted for the next four months of the NSE of India. The plots of these static forecasts are compared with the plot of actuals for the period January 1, 2023 to April 30, 2024 to check the forecasting accuracy. Remarkably, the forecasted values closely follow the actual stock prices in both cases. The low mean absolute percentage error and root mean squared error also validate the fit of the models.

Table 4 presents the ARCH-LM test results for the ARCH effects in the ARIMA residuals. The null hypothesis of no ARCH-GARCH effects is rejected, as the calculated ARCH-LM test probabilities are less than the significance level. Hence, the variances can be forecasted by fitting the



	AC	PAC	Q-Stat	Prob
1	-0.007	-0.007	0.0232	
2	-0.038	-0.038	0.7607	
3	-0.021	-0.021	0.9754	
4	0.004	0.003	0.9846	0.321
5	0.053	0.052	2.4195	0.298
6	0.010	0.011	2.4681	0.481
7	0.034	0.038	3.0393	0.551
8	-0.004	-0.001	3.0476	0.693
9	0.038	0.040	3.7602	0.709
10	-0.029	-0.031	4.1956	0.757
11	-0.056	-0.055	5.7965	0.670
12	0.043	0.038	6.7356	0.665
13	0.002	-0.003	6.7388	0.750
14	-0.020	-0.024	6.9424	0.804
15	0.016	0.021	7.0819	0.852
16	0.005	0.007	7.0955	0.897
17	0.020	0.020	7.3069	0.922
18	0.004	0.007	7.3136	0.948
19	-0.045	-0.043	8.3651	0.937
20	-0.047	-0.045	9.5106	0.923
21	-0.058	-0.069	11.260	0.883
22	0.002	-0.009	11.263	0.915
23	0.029	0.028	11.703	0.926
24	0.047	0.046	12.874	0.913
25	0.071	0.082	15.516	0.839
26	0.024	0.046	15.805	0.863
27	-0.037	-0.024	16.508	0.869
28	0.017	0.027	16.657	0.894
29	0.035	0.027	17.320	0.899
30	0.053	0.038	18.797	0.877
31	-0.015	-0.024	18.910	0.901
32	-0.041	-0.048	19.804	0.899
33	0.025	0.022	20.137	0.913
34	0.012	0.007	20.214	0.931
35	0.011	0.010	20.281	0.946
36	0.022	0.040	20.540	0.955

Figure 3: Correlogram of Residuals of Short-period Data

GARCH model on the residuals of the ARIMA forecasts. Therefore, GARCH (1,1) has been fitted to the forecasts of the resultant residuals of ARIMA forecasting. Table 5 presents the GARCH model estimates. The estimates of the GARCH model provide the effects of both past variances of the forecasted residuals and squares of residuals of those variances on current variances.

Table 4: Heteroscedasticity Test for ARIMA Residuals

Short-period data	F-statistic	4.094	Prob.F(1,492)	0.0436
	Observed R^2	4.076	Prob.Chi square(1)	0.0435
Long-period data	F-statistic	14.155	Prob.F(1,1230)	0.0002
	Observed R^2	14.017	Prob.Chi square(1)	0.0002

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.001	0.001	0.0013	
		2	-0.000	-0.000	0.0013	
		3	-0.013	-0.013	0.2166	
		4	-0.023	-0.023	0.8764	
		5	0.002	0.002	0.8821	
		6	-0.001	-0.001	0.8828	
		7	0.010	0.009	0.9992	0.317
		8	0.014	0.014	1.2562	0.534
		9	0.029	0.029	2.3190	0.509
		10	-0.010	-0.010	2.4540	0.653
		11	0.008	0.009	2.5372	0.771
		12	0.004	0.006	2.5593	0.862
		13	0.003	0.004	2.5678	0.922
		14	0.015	0.015	2.8556	0.943
		15	0.013	0.014	3.0725	0.961
		16	0.018	0.018	3.4959	0.967
		17	0.036	0.036	5.1562	0.923
		18	0.022	0.023	5.7706	0.927
		19	0.004	0.005	5.7890	0.953
		20	-0.023	-0.022	6.4273	0.955
		21	-0.012	-0.010	6.5992	0.968
		22	-0.012	-0.012	6.7765	0.977
		23	-0.008	-0.010	6.8596	0.985
		24	0.017	0.014	7.2398	0.988
		25	0.009	0.006	7.3488	0.992
		26	0.009	0.006	7.4507	0.995
		27	-0.049	-0.050	10.488	0.972
		28	-0.004	-0.003	10.507	0.981
		29	0.038	0.040	12.380	0.964
		30	0.047	0.046	15.155	0.916
		31	0.034	0.032	16.647	0.894
		32	0.015	0.014	16.924	0.911
		33	0.067	0.068	22.583	0.707
		34	0.022	0.025	23.186	0.724

Figure 4: Correlogram of Residuals of Long-period Data

Table 5: GARCH Estimates on ARIMA Residuals

Variable	*Coefficient*	*Standard error*	*p-value*
Residual $(-1)^2$	–0.1036	0.223803	0.6436
GARCH (–1)	0.5856	1.063476	0.5819
Constant	108.7559	238.8774	0.6489

Thus, the estimated variance forecasted by the GARCH model is:

$$\sigma_t^2 = 108.7559 - 0.1036u_{t-1}^2 + 0.5856\sigma_{t-1}^2 \tag{27}$$

Since the forecast of variances by GARCH fitting on short-time data produces highly insignificant results, the stock price variance is neither effected by its past value nor by the residuals of the variance. As none of the residuals has any effect on the forecast of variances, the short-period effect is not observed on the stock price volatility. Therefore, the variances are forecasted again with long-period data and this time fitting the EGARCH model for the time series. The estimated EGARCH results presented in Table 6 show that three out of four coefficients are highly significant.

Table 6: Estimated GARCH Model on ARIMA Residuals

Parameter	*Coefficient*	*Standard error*	*p-value*
α	1.4448	0.0097	0.000
λ	–0.6194	0.0176	0.000
θ	–0.1258	0.0463	0.006
Γ	0.8237	0.0006	0.000

Therefore, the estimated variance forecasting of the EGARCH model is:

$$\log \sigma_t^2 = 1.444 - 0.619 \ln \sigma_{t-1}^2 - 0.126 \frac{u_{t-1}}{\sqrt{\sigma_{t-1}^2}} + 0.824 \left[\frac{|u_{t-1}|}{\sqrt{\alpha_{t-1}^2}} - \sqrt{\frac{2}{\pi}} \right] \quad (28)$$

The estimated coefficient λ, the persistence of conditional volatility, is the least in terms of magnitude out of all the parameters and is highly significant. Hence, the volatility decays at a very short time in the market. Since θ, the asymmetry parameter, is significantly negative, there is a leverage effect implying that positive shocks generate less volatility than negative shocks. The significant positive coefficient γ, measuring the GARCH effect, implies there is a GARCH effect *i.e.* the present variance is effected by its past values.

Figure 5 presents the plot of actual vs forecasted variance by fitting the EGARCH model on stock prices of TATA Steel Limited for the 69 days, from January 1 to April 13, 2024. The plot shows a remarkably close fit of the EGARCH model to the actuals. Especially, the high volatility in the first half of the period gradually decays in the latter half of the time period considered, as the EGARCH model explains. It is to be noted that the actual variance is calculated by taking standard deviations of the actual stock prices for the 69 days and then squaring them to variances, whereas the forecasted variance is generated as a result of the heteroscedasticity model with error variance that includes certain factors like leverage effects, asymmetry, etc. Therefore, in Figure 5 the plot of calculated actual variance shows slightly higher volatility than the plot of the forecasted variance.

It can be observed that the stock volatility is caused by negative shocks as shown by the θ coefficient. The high volatility in stock prices in the month of January is generally attributed to the phenomenon of the "January effect" in the stock market (Thaler, 1987). This happens as investors buy more stocks during the month of December owing to the fall in stock prices that happens during the end of a year and sell them during the last week of January of the next year when stock prices rebound. Also, generally, the total trade volume of stock in December is relatively high in the month of January, leading to relatively more volatility in January. The volatility in stock prices

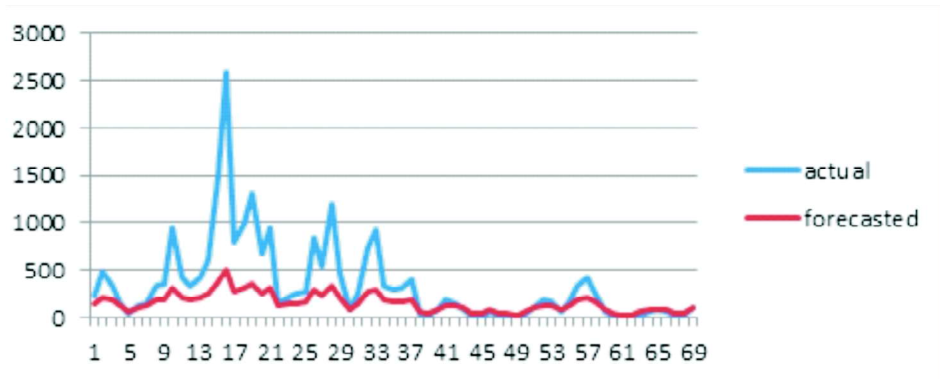


Figure 5: Performance of Actual vs Forecasted Volatility of Stock Prices

that is present during the month of January will not last long as it will decay in the following days, as the coefficient of l parameter reveals. The forecasting performance of the EGARCH model is evaluated by the root mean square, 14.373 for the fitted EGARCH model.

While fitting the ARIMA model for short-period data, the price on the present working day is dependent on prices in the preceding periods along with the errors made in two preceding periods, and for long-period data, the price on the present working day is dependent on longer preceding periods and the errors made in those periods. The parameters in the fitted ARIMA model to short data are all significant, hence the plot for the forecasted values almost overlapped the actual stock prices. In the case of long-period data, all the parameters other than the intercept and the third autoregressive term are highly significant, thus the plot for the forecasted series closely follows the plot of the actual series of stock prices.

However, the residual plots of both time period models show volatility clustering as a result of which the GARCH model has been used to study the variances in the prices in the short data. As the coefficients of the fitted GARCH model are highly insignificant, the ARCH and GARCH terms are quite ineffective in studying the fluctuations in conditional variance. The GARCH model also assumes symmetry in residuals and volatility. In order to capture any possible asymmetry in the stock prices, the EGARCH model is estimated to understand the variance fluctuations for relatively long-period data The parameters that signify symmetry, persistence of volatility and leverage effect are significant, implying that the volatility in the stock prices will die out within a very short span of time. The plot of forecasted variance shows that the volatility in the TATA Steel stock prices that is observed in the month of January does not extend to February. Hence, the volatility in the TATA Steel stock prices is not quite persistent. Thus, the forecasted stock price variance on the basis of a long-period price series is more reliant than the one forecasted on a short-period price series.

CONCLUSION

Generally, an economy is said to be healthy when the stock market is stable. The stock market performance is also considered a barometer of the health of the manufacturing and financial sectors, the important pillars of economic development. An increasing trend in stock market indices implies the growth of the economy and a decreasing trend indicates poor performance. In any stock market, the stock prices are generally volatile and fluctuate over time which are measured by various indices. This paper analyses the effects of long-period and short-period stock price variations on stock prices and using the volatility in residuals *i.e.* error variance, the future volatility in stock prices is forecasted. Empirically, based on the stock prices of TATA Steel Limited in the NSE between January 1, 2021 and April 31, 2024, the short-period and long-period stock prices are used to predict the future stock price and future volatility. First, the ARIMA model has been fitted to forecast the stock prices, and to the resultant residuals, GARCH and EGARCH models have been applied to examine the effects of short-period and long-period price fluctuations on the forecasts of future price variations respectively. The EGARCH fitting shows that the long-period fluctuations have a significant effect on the future stock price volatility relative to the GARCH fitting. The predictability of the EGARCH model has been compared with the actual stock price variations in the next 69 days, from January 1 to April 13, 2024. The econometric forecasts show that the long-period stock price volatility is more reliant than the short-period volatility in forecasting future stock price volatility as well as the stock prices.

REFERENCES

Ahmad, M.H., P.Y. Ping, S.R. Yaziz and N.H. Miswan (2015) "Forecasting Malaysian Gold Prices Using a Hybrid of ARIMA and GJR-GARCH Models", *Applied Mathematical Sciences*, 9, 30, 1491-1501.

Ashik, M.A. and S.K. Kannan (2017) "Forecasting National Stock Price Using ARIMA Model", *Global and Stochastic Analysis*, 4, 1, 77-81.

Banerjee, D. (2014) "Forecasting of Indian Stock Market Using Time-Series ARIMA Model", Conference Proceedings ICBIM-14.

Bollerslev, T. (1986) "Generalised Autoregressive Conditional Heteroskedasticity", *Journal of Econometrics*, 31, 3, 307-327.

Box, G.E.P. and G.M. Jenkins (1976) *Time Series Analysis: Forecasting and Control*, Holden Day, San Francisco.

Brooks, C. (2008) *Introductory Econometrics for Finance*, Cambridge University Press, New York.

Dickey, D.A. and W.A. Fuller (1979) "Distribution of the Estimators for Autoregressive Time Series with a Unit Root", *Journal of the American Statistical Association*, 74, 366, 427-431.

Dritsaki, C. (2015) "Forecasting Real GDP Rate Through Econometric Models: An Empirical Study from Greece", *Journal of International Business and Economics*, 3, 1, 13-19.

Engle, R.F. (1982) "Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of UK Inflation", *Econometrica*, 50, 4, 987-1007.

Epaphra, M. (2017) "Modeling Exchange Rate Volatility: Application of the GARCH and EGARCH Models", *Journal of Mathematical Finance*, 7, 1, 121-143.

Glosten, L.R., R. Jagannathan and D.E. Runkle (1993) "On the Relation Between the Expected Value and the Volatility of the Nominal Excess Return on Stocks", *Journal of Finance*, 48, 5, 1779-1801.

Guha, B. and G. Bandhopadhyaya (2016) "Gold Price Forecasting Using ARIMA Model", *Journal of Advanced Management Science*, 4, 2, 117-121.

Maity, B. and B. Chatterjee (2012) "Forecasting GDP Growth Rates of India: An Empirical Study", *International Journal of Economics and Management Sciences*, 1, 9, 52-58.

Nelson, D.B. and C. Cao (1992) "Inequality Constraints in the Univariate GARCH Model", *Journal of Business and Economic Statistics*, 10, 2,229-235.

Oliveria, R.A., C.W. O'Connor and G.W. Smith (1979) "Short-run Forecasting Models of Beef Prices", *Western Journal of Agricultural Economics*, 4, 1, 45-55.

Su, C. (2010) "Application of EGARCH Model to Estimate Financial Volatility of Daily Returns: The Empirical Case of China", Master Degree Project No.2010:142, Graduate School Master of Science in Finance, School of Business, Economics and Law, University of Gothenburg.

Thaler, R.H. (1987) "Anomalies: The January Effect", *Economic Perspectives*, 1, 1, 197-201.

Yaziz, S.R., N.R. Azizan, R. Zakaria and M.H. Ahmad (2016) "Modelling Gold Price Using ARIMA-TGARCH", *Applied Mathematical Sciences*, 10, 28, 1391-1402.

for citation :

T. Lakshmanasamy (2024). Forecasting Future Stock Price Volatility: Error Variance Estimation. *Journal of Applied Econometrics and Statistics,* Vol. 3, No. 1-2, pp. 97-115.