Multivariate Share Price Forecasting of State Bank of India and Housing Development Finance Corporation Limited Banks

Nithish Kumar M L, Nagarjuna C M and K Varsha

Department of PG studies and Research in Statistics, Mangalore University

Mangalagangotri-574199, Karnataka

ABSTRACT:

Time series analysis is a specific way of analysing a sequence of data points collected over an interval of time. Time is a crucial variable because it shows how the data adjusts over the course of the data points as well as the final results. In this paper, we analysed the daily closing price of shares of State Bank of India (SBI) and Housing Development Finance Corporation Limited (HDFC) that is under National Stock exchange (NSE) dated from 01-01-2016 to 29-01-2021. From the time profile of SBI and HDFC we observe that the data is not stationary and contains trend component. The fluctuations in the two data sets are similar over the time period. Thus, the variations in the stock price over the time period is not much affecting on the accuracy of the fitted model. Since there is cointegration between the share prices under study, we can conclude that there is influence of stock prices of SBI and HDFC banks on each other. After fitting various models and forecasting adjusted close price for the data under study, we see that both multivariate and univariate models have almost similar forecast accuracy. Even though from the findings we conclude that Univariate GARCH family models (i.e., GJR-GARCH(1,1) & E-GARCH (1,1)) may be more suitable for short-term forecasting horizons.

KEYWORDS: ARCH, GARCH, EGARCH, GJR-GARCH,VAR ,Volatility

1.INRODUCTION

Time series analysis is a specific way of analysing a sequence of data points collected over an interval of time. In time series analysis, analysts record data points at consistent intervals over a set period of time rather than just recording the data points intermittently or randomly. However, this type of analysis is not merely the act of collecting data over time. What sets time series data apart from other data is that the analysis can show how variables change over time. The use of time series data for understanding the past and predicting future is a fundamental part of business decisions in every sector of the economy and public service. In time series analysis, variables of interest can be univariate or multivariate. Multivariate processes arise when several related time series are observed simultaneously over time, instead of observing a single series. In this paper we intended to describe the influence and volatility of stock prices of State Bank of India (SBI) and Housing Development Finance Corporation Limited (HDFC) banks on each other by using different financial time series models and compare the forecasting performance. The data contains daily closing price of shares of SBI and HDFC that is listed under National Stock exchange (NSE) dated from 01-01-2016 to 29-01-2021 which is collected from website https://finance.yahoo.com/ .

2.LITERATURE REVIEW

In recent decades, the issue of stock investment, stock market and stock trading are treated with more interest. Most people find it difficult to believe that stock is another viable area to invest in especially at this period of economic doldrums. Usually, the movement of stock market index is influenced by the moves of other stock market indices around the world or in that region, making it a critical topic to monitor over time. As a result, the potential to reliably forecast the future value of stock market indices by taking trade relationships into account is critical. Time series models stand tall in addressing these challenges.

Thus, an extensive work on time series modelling is carried out by many of the researchers. Several models developed to describe the nature of time series data created a bench mark in the literature. Some of the important literatures are reviewed in order to understand the theory behind the modelling technique. A research article by **Stock and Watson (2001)** critically reviews the use of vector auto regressions (VARs) for four tasks: data description, forecasting, structural inference, and policy analysis. The paper begins with a review of VAR analysis, highlighting the differences between reduced-form VARs, recursive VARs and structural VARs. Three variables VAR model that includes the unemployment rate, price inflation and the short-term interest rate is used to show how VAR method is used for the four tasks. They concluded that VARs have proven to be powerful and reliable tools for data description and forecasting, but have been less useful for structural inference and policy analysis. **Runkle, D.E. (2002)** questioned the statistical significance of variance decompositions and impulse response functions for unrestricted vector autoregressions. It suggests that previous authors have failed to provide confidence intervals for variance decompositions and impulse response functions. He developed two methods of computing such confidence intervals **Ahammad Hossain et al. (2015)** carried out a study on Vector Auto Regressive (VAR) models on selected indicators of Dhaka stock exchange (DSE) for the period from June 2004 to July 2013 as the basis on daily scale. The forecast performance of the different VAR models is discussed. **Iberedem and Blessing (2016)** forecasted stocks of the Nigerian banking sector using multivariate time series models. The study involved the stocks from six different banks that were found to be analytically interrelated. They found that the vector autoregressive model of order 1 is more suitable to explain the nature of time series under study. **Jacopo De Stefani (2019)** presented a description of the fundamentals of time series analysis and a review of the state-of-the-art in the domain of multivariate, multiple-step-ahead forecasting. The experimental results show that the proposed strategies by Jacopo De Stefani are a promising alternative to state-of-the-art models, overcoming their limitations in terms of problem size (in case of statistical models) and interpretability (in case of large-scale black-box machine learning models, such as Deep Learning techniques). **Castán-Lascorz et al. (2022)** proposed a new algorithm to predict both univariate and multivariate time series based on a combination of clustering, classification and forecasting techniques. The main goal of the proposed algorithm is first to group windows of time series values with similar patterns by applying a clustering process. The new algorithm has been designed using a flexible framework that allows the model to be generated using any combination of approaches within multiple machine learning techniques. To evaluate the model, several experiments are carried out using different configurations of the clustering, classification and forecasting methods that the model consists of. The results are analyzed and compared to classical prediction models, such as autoregressive, integrated, moving average and Holt-Winters models, to very recent forecasting methods, including deep, long short-term memory neural networks, and to well-known methods in the literature, such as k nearest neighbours, classification and regression trees, as well as random forest.

3. METHODOLOGY

3.1 FORECASTING TECHNIQUES

One of the main objectives of this paper is to study some econometric models available in the literature for modeling the volatility of an asset return. In this work we used Generalised Auto Regressive Conditional Heteroscedastic (GARCH), Exponential GARCH, Glosten Jagannathan and Runkle (GJR) GARCH models under univariate time series modeling and Vector Autoregressive Model (VAR) as that of multivariate case. GARCH is a statistical model that can be used to analyse a number of financial data. The representation of GARCH model is given by,

$$Y_t = \sigma_t \varepsilon_t \; and \; \sigma_t^2 = \alpha_0 + \sum_{i=1}^{p} \alpha_0 Y_{t-i}^2 + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^2$$

where $\{\varepsilon_t\}$ is a sequence of white noise $\alpha_0 > 0, \alpha_i \geq 0 \; and \; \beta_j \geq 0$. Where α_i are ARCH parameters and β_j are GARCH parameters.

The overcome some weakness of the GARCH model in handling financial time series Nelson (1991) proposes the exponential GARCH (EARCH) model. In particular, it allows for asymmetric effects between positive and negative asset by considering the weighted innovation $\varepsilon_t = \theta \varepsilon_t + \gamma[|\varepsilon_t| - E(|\varepsilon_t|)]$. Where θ and γ are real constant. Both ε_t and $[|\varepsilon_t| - E(|\varepsilon_t|)]$ are zero-mean iid sequence with continuous distribution. This model differs from the GARCH variance structure because of the log of the variance.

In a different approach, by considering the fact that the sign and magnitude of the shocks have asymmetric effects on return, Glosten, Jagannathan and Huskie (GJR) introduced GARCH with differing effects of negative and positive shocks taking into account the leverage phenomenon. Thus GJR GARCH model is a generalisation of GARCH model that is appropriate for modelling asymmetric volatility clustering. Specifically, the model posits that the current conditional variance is the sum of linear processes, with past conditional variances. The model captures the asymmetrical nature of a time series by including the indicator function to return data. The function would then produce a value of one if there were a loss and zero if there were profits.

In the VAR model, each variable is modelled as a linear combination of past values of itself and the past values of other variables in the system. VAR modelling technique models the data under study as system of equations with one equation per variable (time series). That is, if you have 5 time series that influence each other, we will have a system of 5 equations. If the series under consideration is stationary, we forecast them by fitting a VAR to the data directly (known as a "VAR in levels"). Otherwise, we take differences of the data in order to make them stationary. Then fit a VAR model (known as a "VAR in differences"). In both cases, the models are estimated by equations using the principle of least squares. The system of equations for a VAR (2) model with three time series namely, Y1, Y2 and Y3 is given by,

$Y_{1,t} = c_1 + \phi_{11,1}Y_{1,t-1} + \phi_{12,1}Y_{2,t-1} + \phi_{13,1}Y_{3,t-1} + \phi_{11,2}Y_{1,t-2} + \phi_{12,2}Y_{2,t-2} + \phi_{13,2}Y_{3,t-2} + \varepsilon_{1,t}$

$Y_{2,t} = c_2 + \phi_{21,1}Y_{1,t-1} + \phi_{22,1}Y_{2,t-1} + \phi_{23,1}Y_{3,t-1} + \phi_{21,2}Y_{1,t-2} + \phi_{22,2}Y_{2,t-2} + \phi_{23,2}Y_{3,t-2} + \varepsilon_{2,t}$

$Y_{3,t} = c_3 + \phi_{31,1}Y_{1,t-1} + \phi_{32,1}Y_{2,t-1} + \phi_{33,1}Y_{3,t-1} + \phi_{31,2}Y_{1,t-2} + \phi_{32,2}Y_{2,t-2} + \phi_{33,2}Y_{3,t-2} + \varepsilon_{2,t}$

3.2 ACCURACY MEASURES

To check the accuracy of the fitted models we used the mean squared error (MSE), mean absolute error (MAE) and mean absolute percentage error (MAPE). It measures the accuracy as a percentage by taking the absolute average (mean) of the ratio of the difference between actual values and predicted values to the actual values. They are scale independent and used to compare forecast performance between different time series.

$$\text{MAPE}= \frac{1}{n}\sum_{i=1}^{n} \left| \frac{y_i-\hat{y}_i}{y_i} \right|$$

Diamantis Koutsandreas (2021) noted that MAPE enables us to evaluate the forecasting accuracy across multiple time series of different scales. Also, it is easy to communicate, especially within businesses and organizations (Kolassa & Martin, 2011). Thus MAPE is the most popular choice to find the accuracy of forecasted series (Fildes & Goodwin, 2007).

4. ANALYSIS AND DISCUSSION

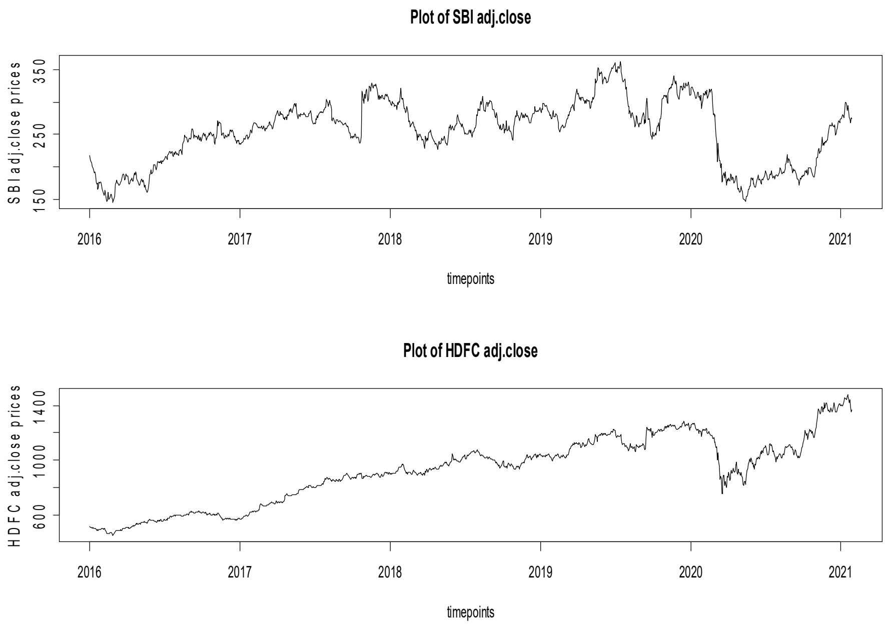

Figure 1: Time profile of stock prices of SBI and HDFC

From Figure 1, we observe that the data is not stationary and contains trend component. The fluctuations in the two data sets are similar over the time period. Further, Mann-Kendall Test confirm that there is presence of monotonic trend in both the series. Then we computed log returns and the summary of log return series is presented in table 1

SUMMARY	HDFC	SBI
MINIMUM	-0.1347	-0.1445
MAXIMUM	0.1097	0.2444
MEAN	0.0007	0.0001
SKEWNESS	-0.5504	0.7211
KURTOSIS	17.3770	15.0610

Table 1: Representing descriptive statistics of log returns of HDFC& SBI

From Table 1, there is excess kurtosis in log returns for both the banks, This can explain that there exist heavier tails in the data and distributed as leptokurtic. We observed that the special features (absence of auto correlation in log return series and presence of autocorrelation in squared log return series) that requires to model volatility of time series along with leptokurtic nature is satisfied by the data sets under study. Further literature on financial time series analysis stress that GARCH(1,1) model is adequate for all practical purposes (Hansen and Lunde, 2004). Thus, we fitted the GARCH model and its extensions with (1,1) parameters. The following table 2 and 3 presents the summary of univariate fitted models.

MODELS	MSE	MAE
GARCH(1,1)	0.0004437001	0.01677733
EGARCH(1,1)	0.0004434944	0.01677603
GJRGARCH(1,1)	0.0004430624	0.01675984

Table 2: Accuracy Measure of the various fitted models for SBI

From table 2, we observe that all the MAE and MSE values of various models is very less that indicates that error rate is too less in these models and are not much differing. Hence by following the results of residual analysis and forecasting performance measures, we conclude that GJRGARCH (1,1) is a suitable model to forecast the SBI adj.close price.

MODELS	MSE	MAE
GARCH(1,1)	0.000225853	0.0122802
EGARCH(1,1)	**0.00022431**	**0.0122206**
GJRGARCH(1,1)	0.000224527	0.0122294

Table 3: Accuracy Measure of the various fitted models for HDFC

From table 3, we observe that all the MAE and MSE values of various models is very less that indicates that error rate is too less in these models and are not much differing. Hence by following the results of residual analysis and forecasting performance measures, we conclude that EGARCH (1,1) is a suitable model to forecast the HDFC adj.close price.

FITTING VAR MODEL AND INTERPRETATION:

Initially, we carried out Engle Granger test to test causality between two datasets under study and observe that p-value is less than the 0.05.Thus we conclude that there is instantaneous causality between SBI and HDFC share price which means that there is interconnection between two share prices. We choose the lag order as 7 to fit VAR model using the information criteria.

Let us consider, Y_1 as the stock price of SBI bank, Y_2 as the stock price of HDFC bank. Then, $\phi_{11,4}\ Y_{1,t-4}$ represents the influence of SBI on itself at 4th lag. $\Phi_{12,4}\ \mathrm{Y}_{2,\,t-4}$ represents the influence of HDFC on SBI at 4th lag and so on. The equations of VAR(7) is given below.

Equation 1: SBI.NS = SBI.NS 1 + HDFC.NS 1 + SBI.NS 2+ HDFC.NS 2 + SBI.NS 3+HDFC.NS 3 + SBI.NS 4+ HDFC.NS 4 + SBI.NS 5+ HDFC.NS 5 + SBI.NS 6+ HDFC.NS 6 +SBI.NS 7+ HDFC.NS 7

Equation 2: Equation 2: HDFC.NS = SBI.NS 1 + HDFC.NS 1 + SBI.NS 2+ HDFC.NS 2 + SBI.NS 3+ HDFC.NS 3 + SBI.NS 4+ HDFC.NS 4 + SBI.NS 5+ HDFC.NS 5 + SBI.NS 6+ HDFC.NS 6 + SBI.NS 7+ HDFC.NS 7

FORECAST:

Further, we forecasted the closing price of HDFC and SBI stock price for next 11 days using univariate and multivariate models. Following table 4 and 5 represent the forecasted and actual values from 01-02-2021 to 12-02-2021.

Date	Forecasted value using EGRCH	Forecasted value using VAR	Actual value
01-02-2021	1369.961	1372.66	1453.486
02-02-2021	1371.319	1374.143	1535.966
03-02-2021	1372.672	1364.942	1549.992
04-02-2021	1374.018	1369.894	1554.224
05-02-2021	1375.359	1372.398	1579.433
08-02-2021	1376.695	1370.191	1579.962
09-02-2021	1378.026	1369.655	1586.458
10-02-2021	1379.352	1368.946	1556.832
11-02-2021	1380.675	1370.562	1547.58
12-02-2021	1381.993	1369.25	1557.029

Table 4: Forecasted closing price of HDFC bank

Date	Forecasted value using GJRGRCH	Forecasted value using VAR	Actual value
01-02-2021	275.315	274.837	303.127
02-02-2021	275.413	274.653	324.981
03-02-2021	275.52	272.835	327.762
04-02-2021	275.633	274.201	346.445
05-02-2021	275.753	272.763	383.519
08-02-2021	275.879	272	387.372
09-02-2021	276.01	272.406	385.177
10-02-2021	276.147	272.298	382.689
11-02-2021	276.287	272.67	380.641
12-02-2021	276.432	272.196	383.567

Table 5: Forecasted Closing price of SBI

The above tables 4 and 5 shows that the forecasts for adj.close price for the SBI and HDFC data under examination are nearly same for both VAR model and the Univariate GARCH family models (i.e., GJR GARCH(1,1) & E GARCH(1,1)).

	MAPE	
SBI	VAR	0.3206
	GJR-GARCH	0.3069
HDFC	VAR	0.1307
	E-GARCH	0.1307

Table 6: Representing the accuracy of forecasted models based on MAPE

From table 6, we notice that both VAR and Univariate GARCH family models (i.e. GARCH(1,1) & EGARCH(1,1)) have almost similar forecast accuracy.

5. CONCLUSIONS:

From the time profile of SBI & HDFC we observe that the data is not stationary and contains trend component. The fluctuations in the two data sets are similar over the time period. Thus, the variations in the stock price over the time period is not much affecting on the accuracy of the fitted model. Since there is cointegration between the share prices under study, we can conclude that there is instantaneous influence of stock prices of SBI and HDFC banks on each other.

After fitting various models and forecasting adj.close price for the data under study(SBI & HDFC data) ,we see that both multivariate and univariate models have almost similar forecast accuracy.

Strictly speaking by taking account the above findings we conclude that Univariate GARCH family models(i.e., GJR-GARCH(1,1) & E-GARCH (1,1)) may be more suitable for short-term forecasting horizons.

Although it is very important to handle the relationships between the different stock prices under study than ignoring it, carrying out univariate and multivariate time series analysis for prediction may end up giving similar results when the relationship between the series are fragile.

In summary, even if univariate and multivariate models have similar forecast accuracy, there may be other factors to be consider when choosing between them, such as data complexity, forecast horizon, model interpretability, flexibility etc. It is important to carefully evaluate the specific requirements and constraints of the forecasting task and consider all relevant factors when making a decision.

6. REFERENCES

- Alev, D. A. & Seyma, C.C. (2015). Comparison of Prediction Performances of Artificial Neural Network (ANN) and Vector Autoregressive (VAR) Models by Using the Macroeconomic Variables of Gold Prices, Procedia Economics and Finance, Vol. 30, 3-14. https://doi.org/10.1016/S2212-5671(15)01249-6.

- Diamantis, K., Evangelos, S. & Fotios, P.(2022). On the Selection of Forecasting Accuracy Measures, Journal of the Operational Research Society,Vol.73,937-954. https://doi.org/10.1080/01605682.2021.1892464

- Ginika, N., Odokp, P., Jeremiah,A. & Esuabana, I. (2016). Modelling and Adequacy of Vector Autoregressive Model.

- Hossain, A., Kamruzzaman, M. & Ali, M. (2015). Vector Autoregressive (VAR) Modeling and Projection of DSE. Chinese Business Review, Vol.14, 273-289. doi:10.17265/1537-1506/2015.06.001.

- Iwok, I & Okoro, B. (2016). Forecasting Stocks with Multivariate Time Series Models, International Journal of Mathematics and Statistics

Invention, Vol.4, 32-44.

- Johann, P. & Stephen F. W. (2003).Univariate versus Multivariate Time Series Forecasting: An Application to International Tourism Demand, International Journal of Forecasting, Vol.19(3), 435-451. doi:org/10.1016/S0169-2070(02)00057-2.

- Patrick, A., Qun, M. Frank, M., Sanfilippo, D. B.(2015).A Comparison of Multivariate and Univariate Time Series Approaches to Modelling and Forecasting emergency Department Demand in Western Australia, Journal of Biomedical Informatics,Vol.57,62-73, https://doi.org/10.1016/j.jbi.2015.06.022.

- Rucy S.Tsay(2009).Analysis of Financial Time Series. 2nd Edition. Wiley Series in Probability and Statistics,ISBN 978-81-265-2369-6.

- Vievien, A., Djara, D. Dhita , D., Hananti, H., Qisthi, N. & Rosmanah. (2022). Prediction of Export and Import in Indonesia Using Vector Autoregressive Integrated (VARI), doi:105. 10.28919/jmcs/71.