2020

Modeling and Forecasting Selected Indicators of Dhaka Stock Exchange

Hossain, Ahammad

University of Rajshahi

Modeling and Forecasting Selected Indicators of Dhaka Stock Exchange

A Dissertation

Submitted to the Institute of Bangladesh Studies (IBS), University of Rajshahi in Partial Fulfillment of the Requirements for the Award of the Degree of Doctor of Philosophy in Statistics

Supervisor
Professor Dr. Md. Ayub Ali
Department of Statistics
University of Rajshahi
Rajshahi-6205
Bangladesh

Researcher
Ahammad Hossain
Registration No. 0410
Session: 2012-2013
Institute of Bangladesh Studies
University of Rajshahi
Rajshahi-6205, Bangladesh

INSTITUTE OF BANGLADESH STUDIES

UNIVERSITY OF RAJSHAHI

Rajshahi-6205, Bangladesh

December 2020

Dedicated to the Soul

of

My Beloved Parents

Declaration

I hereby declare that the study entitled "**Modeling and Forecasting Selected Indicators of Dhaka Stock Exchange**" submitted to the Institute of Bangladesh Studies (IBS), University of Rajshahi in partial fulfillment of the requirements for the award of the Degree of Doctor of Philosophy in Statistics, is the result of my research, except where otherwise referenced or acknowledged under the supervision of Dr. Md. Ayub Ali, Professor, Department of Statistics, University of Rajshahi, Bangladesh, and this dissertation or any part of the same has not been submitted for qualification at any other university or institution.

Ahammad Hossain
PhD Fellow
Session: 2012-2013
Registration Number: 0410
Institute of Bangladesh Studies (IBS)
University of Rajshahi
Rajshahi-6205
Bangladesh.

Date: December 5, 2020

Certification

This is to certify that the dissertation entitled "**Modeling and Forecasting Selected Indicators of Dhaka Stock Exchange**" submitted by Mr. Ahammad Hossain, a PhD Fellow, Registration Number: 0410, Session: 2012-2013, Institute of Bangladesh Studies (IBS), University of Rajshahi in partial fulfillment of the requirements for the award of the Degree of Doctor of Philosophy in Statistics. It is also certified that the research work embodied in this dissertation is original and carried out by him under my supervision. To the best of my knowledge, any part of this dissertation has not been submitted elsewhere for awarding any other degree. This study has proposed sophisticated models to forecast stock indices of the Dhaka Stock Exchange. He proposed highly stable models that behave almost fitted accurately and can assess the risk for the investors. That's why; I believe that this study will be served in policy implication and decision making of the investors as well as other financial sectors. From this dissertation, one paper in Chinese Business Review, one paper in Economics World, and one paper in Journal of Statistical Science and Application have been published and one paper has been presented in an International Conference on Statistical Data Mining for Bioinformatics, Health, Agriculture and Environments, Department of Statistics, University of Rajshahi. Also, one paper is under construction for possible publication. It is also certified that I have gone through the draft and final version of the dissertation and approved it for submission.

Dr. Md. Ayub Ali
Professor
Department of Statistics
University of Rajshahi
Rajshahi-6205
Bangladesh.
Date: December 5, 2020

Acknowledgments

All the praises and thanks are due to the Almighty Allah who has created us and given me strength, course, patience, and ability to make this dissertation possible, and for surrounding me with people who helped me accomplish this goal, therefore, they deserve special recognition.

I wish to express my profound and heartfelt gratitude to my reverend supervisor, Dr. Md. Ayub Ali, Professor, Department of Statistics, University of Rajshahi for his cordial supervision, invaluable guidance, constant encouragement, his wide knowledge, and helpful conversation throughout the progress of this dissertation. I am highly grateful to him for allowing me to work under his supervision.

I would like to express my sincere gratitude to the honorable director of the Institute of Bangladesh Studies (IBS), Professor Dr. Jakir Hossain for his kindness and invaluable suggestions to improve my dissertation. I am very grateful to him.

I would like to express my sincere gratitude to my honorable teacher, Dr. Md. Kamruzzaman, Associate Professor, IBS, University of Rajshahi for his valuable suggestions, helping and inspiring me at every stage of this study. I am very grateful to him for his continuous support and encouragement to carry out this dissertation.

I also would like to thank faculty members and staff at IBS, University of Rajshahi for their help and support during this study.

I am most grateful to my parents to whom I owe my deepest gratitude for inspiring me and making me appreciate the value of education.

In addition, my gratitude goes to the Ministry of science and technology, Govt. of the People's Republic of Bangladesh for providing me with a research stipend.

Finally, I owe a debt of gratitude to my wife Ashifa Khatun Konika, daughter Ayra Adiba, elder sister Mst. Shakhina Khatun and younger brother Mohammad Hossain for their tremendous patience and invaluable support throughout the research.

Ahammad Hossain

Rajshahi, Bangladesh

December 5, 2020

Preface

Dhaka Stock Exchange (DSE) is functioning as a strong mechanism for industrialization and economic growth in Bangladesh. The development of the stock market is essential for capital accumulation, proficient distribution of resources, and elevation of economic progression. An animated stock market is expected to support an economy to be vigorous, but two major cataclysms in the history of the stock market of Bangladesh during one and half decades do not show the presence of a vibrant market. However, these catastrophes demonstrate an extremely risky and unstable capital market. These cataclysms have also stunned the whole country, as millions of stockholders became insolvent in a very short period of time (Islam, 2011). The DSE General Index (DGI) became the highest ever in 2010, whereas the investors found it as the lowest ever in the first quarter of 2011 (DSE General Index Data, 2012).

Stock market modeling and forecasting is a very tough job, but it is becoming popular among scholars for theoretical and technical reasons in the world. Imperfection models can be a threat to the investors and the researchers due to the gap of theoretical knowledge. Stock prices are evaluated in the marketplace, where the stock sellers meet the buyers' demand. The share prices do not follow a definite model. We know little about the forces of stock indices that accelerate the share price unexpectedly up or down. These forces are categorized into three types, e.g., fundamental factors, technical factors, and market sentiment. In a well-organized market, fundamental factors like earnings per share (EPS) and a valuation multiple (P/E ratio) evaluate stock prices. Technical factors, the mixture of exterior surroundings, control the supply of demand of a company's stock. Some of these eventually affect the fundamentals. Technical factors contain demographics, trends, liquidity, the economic strength of the market and peers, substitutes, incidental transactions, inflation, etc. The individual and collective mindset of the stockholders defines the market sentiment. It is conceivably the most troublesome category. The market sentiment of DSE is analyzed from the literature of institutional bodies like DSE, Bangladesh Securities and Exchange Commission (BSEC), Bangladesh Bank, World Bank, etc. Market sentiment is often stubborn, unfair, and subjective. A shareholder can make a solid judgment to buy or sell shares through the projections on market sentiment. Shareholders expect the attention of institutional investors on fundamental factors.

Stock market modeling has become an essential issue in the finance literature for volatility forecasting. Although a lot of models have been proposed and key parameters on forecasting volatility have been implemented, finance research has not reached a solid consensus regarding this issue. This dissertation works with the unending debate of using various families of models from the existing pool of models as well as proposes and explores the further crucial parameters in improving the accuracy of forecasting. The consequence of accurate volatility forecasting is supreme for a well-functioning economy. Knowledge, understanding, and the ability to estimate proxy volatility are the determining factors for the individual and institutional investors, researchers, and academics. The most reasonable time series indicators of DSE from fundamental factors and technical factors have been selected and various statistical models have been proposed to forecast the volatility of DSE. The volatility modeling and forecasting of DSE indices using different scales of time series are essential for numerous areas of financial analysis in Bangladesh. Thus, volatility modeling and forecasting successfully affect portfolio selection, risk management, option pricing, and monetary policy-making.

Abstract

This study was an attempt to build a suitable time series model to forecast selected indicators of Dhaka Stock Exchange (DSE). Time series data of STR, IMC, TEC during 1990-2012 in annual scale; DGI, GDP, GNI, GS, GI, DIR and GFI during 2005-2012 in annual scale; capital, volume, value, trade, DGI during 2004- 2013 in monthly scale; and DSEX, DSES, and DSE30 indices during 2014-2018 in monthly scale were used for modeling and forecasting purposes. The data were collected from the World Bank and DSE websites. Exploratory Data Analysis (EDA) was used to uncover the hidden information carried out through the observed data. The time series plot showed that DSE indicators had a rightly upward trend over time but non-seasonality was present in the series.

Cobb-Douglas (CD) functional regression of STR on IMC and TEC was estimated. There was a negative STR trend during the period from 1990 to 2012. There was no multicollinearity problem among the regressors in the CD regression model. The estimated residuals of the CD model satisfied that the model was free from the problem of outliers and also confirmed the normality condition. To investigate the indirect and long-run impact on the portfolios of DSE prices, the multiple log-linear regression model was estimated considering the DGI as the dependent variable and the macroeconomic indicators like GDP, GNI, GS, GI, DIR, and GFI, respectively as the independent variables. A negative DGI trend (α = -44.936) was found during the period 2005 to 2012. Multicollinearity, normality, and outliers were checked for this model too. For the multicollinearity problem, the multiple linear regression model was re-estimated by dropping GNI due to very severe multicollinearity and for severe/moderate multicollinearity, standardized GDP, standardized GS, and standardized GFI were used as the explanatory variables. A positive DGI trend during the period 2005 to 2012 was found.

To fit proper ARIMA, VAR, and ARIMA with GARCH family models, the stationary property of the series was confirmed. A suitable VAR(2) model was

finally selected on the basis of AIC and BIC model selection criteria and then estimated with the stock trade, invested stock capital, stock volume, current market value, and DGI on a monthly scale. Johansen cointegration test results suggested that none of the series of the estimating VAR models were cointegrated. The normality test of the estimated residuals of the VAR(2) model suggested that the residuals were a lack of normality. The estimated residuals from VAR(2) model rejected the null hypothesis of no ARCH effects. Granger causality test results suggested that there were bivariate causal relationships among the variables of estimated VAR models. The auto ARIMA models of capital, DGI, value, volume, and trade data series were estimated using auto.arima( ) function of R Package ‘forecast’. The performance of the estimated VAR(2) model was compared with different univariate ARIMA(1,1,1) models. The estimated VAR(2) model performed well than the univariate ARIMA(1,1,1) model of market capital, DGI, and volume data series of DSE. Unfortunately, the DGI count was suspended after July 31, 2013. So, the forecasting of DGI and its associated variables are less valuable for near future analysis of DSE portfolios. Due to the suspension of the DGI and also for the demand of the forecasting of DSE current indicators, univariate ARIMA, ARIMA with GARCH family, ANN, and SVM models were estimated using DSEX, DSES, and DSE30 indices for forecasting purposes. The total model selection was conducted based on training and test performance. Firstly, the stationary conditions of DSEX, DSES, and DSE30 indices were checked. The auto ARIMA models of DSEX, DSES, and DSE30 indices were selected and then estimated using auto.arima( ) function of R Package ‘forecast’. After that, the best-performed model from the GARCH family was selected using the minimum value of AIC and BIC. Finally, the finite mixtures of ARIMA with GARCH family models were established for forecasting purposes. The normality of residuals for each model was checked. Similarly, the best-performed ANN and SVM models were established for each of the DSEX, DSES, and DSE30 indices of DSE.

Finally, this study established that ARIMA(1,0,0) with EGARCH(1,1,2) for forecasting DSEX index, ANN (MLP 3-7-1 net) for forecasting DSES index, and ARIMA(1,0,1) with ARCH(2) for forecasting DSE30 index were found as the most reliable models. The forecasting was conducted from January 2019 to December 2025. The forecasting results of the study may help BSEC, individual and institutional investors, industry owners, stakeholders, and above all the Government of Bangladesh to take appropriate actions for building an efficient and sustainable stock market in Bangladesh.

Share index accounts for the changes in stock prices that are generally connected with the changes in the market stipulation. The shareholders may consider it as a benchmark to see the share market condition in reference to earnings or dividend per share. Again the market condition of every company depends on the financial condition of the country. Thus, DSE indicators forecasting is crucial to clarify the stability of the economic condition of a country. The modeling and forecasting concepts utilized in this dissertation are useful for the shareholders or researchers to work out the long-run value of the share index and thereby taking decisions for investment.

Publications from This Dissertation

The following is a list of publications from this dissertation-

1) **Hossain. A.,** Ali, M.A. and Ali, ABM S. (2012). Time-Varying Volatility Analysis and Forecasting Volume Data in Dhaka Stock Exchange. In peer-reviewed Proceedings- *International Conference on Statistical Data Mining for Bioinformatics, Health, Agriculture and Environments*, 21-24 December. Department of Statistics, University of Rajshahi, ISBN: 978-984-33-5876-9.

2) **Hossain, A**., Kamruzzaman, M., and Ali, M.A. (2015). Vector Autoregressive (VAR) Modeling and Projection of DSE. *Chinese Business Review*, Vol. 14, No. 6. doi: 10.17265/1537-1506/2015.06.001

3) **Hossain, A**., Kamruzzaman, M., and Ali, M.A. (2015). ARIMA with GARCH Family Modeling and Projection on Share Volume of DSE. *Economics World*, Vol. 3, No. 7-8. doi:10.17265/2328-7144/2015.0708.003

4) **Hossain. A**., Kamruzzaman, M., and Ali, M.A. (2016). Econometric Investigation of DSE's Portfolios through Selective Micro and Macro Economic Indicators. *Journal of Statistical Science and Application*, Vol. 4, No. 05-06. doi: 10.17265/2328-224X/2016.0506.003

5) **Hossain. A**., Kamruzzaman, M., and Ali, M.A. (2021). Modeling and Forecasting DSEX, DSES and DSE30 Indices of DSE (under construction).

Contents

Captions	Page No.

Captions	Page No.

Captions	**Page No.**

Captions	Page No.

Captions	**Page No.**

Captions	Page No.

Captions	Page No.

List of Tables

List of Figures

List of Abbreviations

ACF	Autocorrelation Function
ADF	Augmented Dickey-Fuller Test
AIC	Akaike Information Criterion
ANN	Artificial Neural Network
AR	Autoregressive
ARCH	Autoregressive Conditional Heteroskedasticity
ARIMA	Autoregressive Integrated Moving Average
ARMA	Autoregressive Moving Average
BFGS Algorithm	Broyden Fletcher Goldfarb Shanno Algorithm
BIC	Bayesian Information Criterion
BSEC	Bangladesh Securities and Exchange Commission
CAPM	Capital Asset Pricing Model
CD	Cobb-Douglas
CPI	Consumer Price Index
DF	Dickey-Fuller Test
DGI	DSE General Index
DIR	Deposit Interest Rate
DSE	Dhaka Stock Exchange
DSE30 Index	DSE 30 Leading Companies Index
DSES Index	DSEX Shariah Index
DSEX Index	DSE Broad Index
EDA	Exploratory Data Analysis
EGARCH	Exponential Generalized Autoregressive Conditional Heteroskedasticity
EMH	Efficient Market Hypothesis
EPSEAL	East Pakistan Stock Exchange Association Limited
FIGARCH	Fractionally Integrated GARCH
GARCH	Generalized Autoregressive Conditional Heteroskedasticity
GDP	Gross Domestic Product
GED	Generalized Error Distribution

GFI	Gross Foreign Investment
GI	Gross Inflation
GNI	Gross National Income
GNP	Gross National Product
GS	Gross Saving
IMC	Invested Market Capital
IOSCO	International Organization of Securities Commissions
JB Test	Jarque-Bera Test
KPSS Tests	Kwiatkowski Phillips Schmidt Shin Tests
LB Test	Ljung Box Test
LM Algorithm	Levenberg Marquardt Algorithm
LS	Least Squares
MA	Moving Average
MLP	Multilayer Perceptron
MSE	Mean Square Error
OLS	Ordinary Least Squares
PACF	Partial Autocorrelation Function
PARCH	Power ARCH
P/E	Price Earnings Ratio
PP Test	Phillips Perron Test
RBF	Radial Basis Function
RBFT Algorithm	Redundant Byzantine Fault Tolerance Algorithm
RMSE	Root Mean Square Error
SC	Schwarz Criterion
STR	Stock Traded/Turnover Ratio
SVM	Support Vector Machines
TARCH	Threshold ARCH
TEC	Total Enlisted Company
TV	Tolerance Value
VEC	Vector Error Correction
VIF	Variance Inflation Factor
WB	World Bank

Chapter One

Thesis Preliminary

1. Introduction

The stock market functions as a powerful catalyst for a country's commercial and financial development. It is the engine of economic growth. It plays an important role in capital gathering, the allocation of resources, and the attainment of economic progress. A stock market mediates between the surplus and deficit funds of an economy and accelerates savings into investments. It leads to the economic growth of a country by providing finance to the enlisted companies, industries, and firms. Bangladesh, a non-industrial nation with fragile infrastructure and poor investment, requires a coordinated and well-functioning stock market in order to bring about industrialization and economic growth. A major stock market in Bangladesh is the Dhaka Stock Exchange (DSE). It introduced an essential transformation in the last decades. As a result of the appropriation of worldwide quality exchange, settlement components, as well as reductions in exchange expenses, both domestic and foreign investors have become increasingly optimistic, which has led to an impressive increase in market volume and liquidity. DSE maintains an administrative system, current market framework, exclusion of barriers to foreign equity investments, glowing allocation of and utilization of domestic capitals, and market transparency. DSE reforms its indices' liquidity and size. It also notes the number of traded companies, volume, and value of trading, invested capital, current market value, new equity issues, etc. DSE has better efficiency based on all of these factors. Despite this improvement, the DSE shows greater volatility, which harms informational efficiency. Although the DSE experienced volatility from the beginning, it reached its highest level in May 2010 before ultimately falling. This led to a loss of confidence among stockholders (Islam, 2011). The regulators introduced an automated transaction scheme with a circuit breaker that was expected to improve informational efficiency. As a result, both stockholders and companies want to trade at fair prices. Stockholders and companies deserve an efficient stock market. However, the efficiency of the capital market is still questionable. There is one school of thought that believes that the stock market is efficient, while another holds the opposite view. The researchers have developed three types of market hypotheses based on the efficiency of the stock market. Stock prices and returns that are unpredictable indicate a weakly efficient market (Fama, 1991). This market sticks to a random walk model. If the current share prices reflect all public information, then the stock market is deemed semi-efficient. Finally, if share prices disclose all secret and public data, then the market is considered to be strongly efficient. Conventional specialists agree that the share price reflects the actual market value of upcoming dividends. It makes the stock market more efficient. Thus, market investors make decisions on whether or not they want to buy or sell stocks. Financial institutions and banks invest in those portfolios which yield the highest returns. The majority of such investments are made by institutional investors, such as public and private commercial banks and insurance companies. Using monetary reports, financial media, databases, and the internet to analyze a company's financial performance, financial managers, investors, and stockholders convey their decisions. Stock turnover ratio analysis provides individuals and institutions with valuable information about profitability, competence, and risk. Turnover ratio, earnings per share, net asset value, etc. of stocks traded are more closely linked to the portfolios of shareholders. In order for a country to grow economically, it needs a sound and safe stock market. A sound and safe stock market can provide capital to industry owners from their shareholders for productive investment.

1.1 Statement of the Problem

Bangladesh's major stock exchange is the Dhaka Stock Exchange (DSE). Stock prices and other market assets play an important role in determining the economic activity of a country. DSE has also taken significant steps to develop a stable capital market. Modeling and forecasting stock indicators have become popular among researchers for theoretical and technical purposes. The renowned statistician George Box theorized that "All models are wrong, but some are useful" (Box, 1976). A runner can reach the destination by both walking and cycling, but the time consumption may differ. Simply, cycling is faster than walking. Similarly, all the models cannot forecast stock prices properly, but some can. From this point of view, maximum estimation of the forecast returns may be carried out by fitting proper statistical models with applying minimum resources like data and time. A wide variety of private and public economic agents are interested in investing and making financial policies based on the volatility of the stock market indices, stock traded values, market capital, stock volume, and interest rates. A forecasting model that can effectively predict these stock indicators plays a critical role in decision-making. To make predictions, researchers use univariate models like the Autoregressive Integrated Moving Average (ARIMA) and the Generalized Autoregressive Conditional Heteroskedasticity (GARCH) family. In addition, it is important to analyze the interaction between the stock indicators variables in a multivariate framework. Multiple linear regression models using Cobb-Douglas (CD) production function, ARIMA models for mean stabilization, GARCH family models for volatility stabilization, Vector Autoregressive (VAR) models, and Artificial Neural Network (ANN) models are applied to answer the questions presented. To analyze the long term impact of DSE portfolios, the macro economic variables such as Gross National Income (GNI), Gross Domestic Product (GDP), Gross savings (GS), Gross Inflation (GI), Deposit Interest Rate (DIR), and Gross Foreign Investment (GFI) are studied. Stock market modeling and forecasting can be difficult, but it can also mislead the researcher if the series is non-stationary. So, more concentration needs to be given to make the series stationary. In forecasting, models with the lowest residual can perform better than other models. The residuals analysis is also used to check the stock market's dynamics.

1.2 Research Questions

Stock indicator modeling and forecasting can be challenging due to their volatility. In addition, the researchers are trying to come up with new models every day, such as linear, nonlinear, univariate, multivariate, etc., to predict the future trend of stock indicators with handling the time series data. In many cases, the forecasting does not fit the out-of-sample period. The research questions for this study are as follows:

a) How prevalent are certain indicators on DSE portfolios?
b) Which types of models are appropriate for the selected indicators of DSE?
c) How well do the proposed models perform in training and testing?
d) How efficient or stable are the proposed models?
e) Finally, forecast selected indicators out-of-sample using the proposed model.

1.3 Bangladesh Securities and Exchange Commission (BSEC)

In the economy of a country as a whole, the stock market has an important role to play in securing long-term financing. The industrialization and financial progress of a country are based on a reasonable, productive, and straightforward stock market. Bangladesh Securities and Exchange Commission (BSEC) was instituted as a regulatory body on June 08, 1993, following the passage of the Bangladesh Securities and Exchange Commission Act, 1993 with the following mission (BSEC, 2014).

- Keeping the right of protection among stockholders;
- Developing sustainable stock markets; and
- Regulate the rules of securities.

The BSEC defines the rules for regulating stock market activities. It also controls whether stock issuers and market intermediaries are fulfilling their legal obligations. The government appoints a chairman and four full-time commissioners. The chairman is the chief executive of the commission.

1.4 Dhaka Stock Exchange (DSE)

The DSE was founded on 28 April, 1954. It was named as the East Pakistan Stock Exchange Association Limited (EPSEAL). But the trading was started in 1956 with 196 listed companies. EPSEAL had a total of Taka four billion as the paid-up capital (Chowdhury, 1994). It was renamed as Dhaka Stock Exchange Limited on 23 June, 1962. The trading of DSE was postponed due to the liberation war in 1971. It restarted operation in 1976 with only 9 listed companies with a paid-up capital of Taka 0.138 billion and market capitalization of Taka 0.147 billion, which was 0.138% of GDP (Chowdhury, 1994). DSE was enlisted as a Public Limited Company and its operational activities are controlled by its articles of association and its own rules, regulations, and by-laws along with the Securities and Exchange Ordinance 1969, the Company Act, 1994; the Securities and Exchange Commission Act, 1993 (DSE Ltd. Web portal, 2019). In order to run the market efficiently, various policies have been adopted by DSE and BSEC. DSE indices, market capital, the number of enlisted companies, the stock traded value, and stock turnover ratio, etc. are increased annually.

1.5 Selected Indicators of DSE

Secondary data are used for this study. Secondary data like Invested Market Capital (IMC) (USD) and the number of Total Enlisted Company (TEC), Stock Traded/Turnover Ratio (STR), Gross Domestic Product (GDP), Gross National Income (GNI), Gross Saving (GS), Gross Inflation (GI), Deposit Interest Rate (DIR), and Gross Foreign Investment (GFI) were collected from the World Bank website and DSE General Index (DGI) (All share prices of A, B, G & N categories of the portfolios of DSE), DSEX Index (DSE Broad Index), DSES Index, and DSE30 Index were collected from DSE website.

1.5.1 Invested Market Capital (IMC)

Invested Market Capital is the sum of the market value of outstanding shares of a company. To calculate a company's stock capitalization, multiply its outstanding shares by its current market price per share (World Bank WDI, 2017).

1.5.2 Total Enlisted Company (TEC)

Total Enlisted Company refers to the number of companies that have the legal right to trade individual portfolios of companies on the stock market (World Bank WDI, 2017).

1.5.3 Stock Traded/Turnover Ratio (STR)

Stock Traded/Turnover ratio stands for the overall value of stocks trading throughout the period divided by the average market capitalization. Average market capitalization is the mean of the end period and current period values (World Bank WDI, 2017).

1.5.4 DSE General Index (DGI)

DSE General Index is the closing price index of all share prices of A, B, G & N categories with the respective market capital of the portfolios of DSE which is calculated according to the Index algorithm from the International Organization of Securities Commissions (IOSCO) as follows:

DSE General Index (DGI) = $\frac{\text{Yesterday's Closing Index} \times \text{Closing Market Capital}}{\text{Opening Marker Capital}}$

Closing Market Capital =$\sum(\text{Closing Price} \times \text{Total no. of Indexed Shares})$

The DSE General Index (DGI) count was suspended on 31 July in 2013 (DSE Ltd. Web Portal, 2019).

1.5.5 DSEX Index

The Dhaka Stock Exchange Limited computes the DSE broad index which is known as the DSEX index. The DSEX index covers around 97% of the total equity market capitalization. Financial viability is not essential for index membership. Sector-wise diversification rules do not function on the DSEX index. It is generated on free float shares by following S & P Dow Jones indices methodology from 28 January 2013. The DSEX index excludes mutual funds, bonds, and debentures. It had a base value of 2951.91 on 17 January in 2008. It began with 4055.91 points on 28 January 2013 and it remained at 4583.11 points at the end of the financial year 2014-2015. It touched 4507.58 points at the end of the financial year 2015-2016. It fell by 75.53 points or 1.65% in the financial year 2015-2016, however, it reached the highest level at 4873.96 points on 5 August in 2015 and the lowest level at 4171.41 points on 2 May 2016 (DSE Ltd. Web Portal, 2019).

1.5.6 DSEX Shariah Index (DSES)

The DSEX Shariah Index (DSES) was introduced on 20 January in 2014. This index was developed for Islamic Shariah compliant companies in the stock market. The Islamic Shariah board developed the Shariah index by following Standard & Poor’s methodology. In the year 2011, the base value of the DSES index was 1000 points. It was 941.28 points on 20 January in 2014, and it reached 1122.03 points at the end of 2014-2015. At the end of the financial year 2015-2016, it reached 1110.83 points by decreasing 1%. In the financial year 2015-2016, it stood at

1207.92 points as the peak on 4 August in 2015 and it was 1020.02 points as lowest on 2 May in 2016 (DSE Ltd. Web Portal, 2019).

1.5.7 DSE30 Index

The DSE30 index, an investable index of the exchange, was developed for the leading thirty companies. It accounts for about 51% of the total stock market capital. A float-adjusted market capital over Taka 500 million, as on the rebalancing reference date, is the precondition of suitable stocks. The daily average values of three months of stocks need to touch Taka 5 million as on the rebalancing reference date. Institutional stockholders with enough constituents in the index get the advantages of diminished liquidity criteria as minimum Taka 3 million. The institutional stockholders having below Taka 3 million at every semi-yearly rebalancing need to meet additional eligible criteria. A positive net compensation of thirty leading companies over the recent year must be profitable on a rebalancing reference date by evaluating the recent four quarters of compensation reports. DSE keeps a record of all DSE30 indices based on the industry classification framework. The quantity of constituents in banks, financial institutions, insurance sectors, real estate, sub-sector of service and real estate part, pharmaceuticals, fuel, and power is consolidated for the DSE30 index (DSE Ltd. Web Portal, 2019).

1.5.8 Gross Domestic Product (GDP)

Gross Domestic Product (GDP) refers to the monetary value of all finished goods and services produced within a country's borders in a specific period (Investopedia, 2018). It covers all private and public consumption, government expenditures, all investments, and net exports that occur within a defined territory. GDP is calculated using the following formula:

$$GDP = C + G + I + NX$$

where, C is equal to all private consumption, or consumer spending in a nation's economy, G is the sum of government spending, I is the sum of all the country's investment including businesses capital expenditures and NX is the nation's total net exports calculated as total exports minus total imports. NX can be expressed as NX = Exports − Imports.

1.5.9 Gross National Income (GNI)

Gross National Income refers to GDP minus net taxes on production and imports, minus compensation of workers and property income payable to the rest of the world plus the corresponding items received from the rest of the world. An additional approach to measuring GNI at market prices is the overall value of the balances of gross primary incomes for all sectors. Gross national income is an alternative process of calculating GNP. GNI is the sum of a nation's gross domestic product and the net income received from overseas (Investopedia, 2018).

1.5.10 Gross Saving (GS)

Gross Saving (GS) is disposable income less consumption. It can be calculated for each institutional sector and the total economy (World Bank WDI, 2017).

1.5.11 Gross Inflation (GI)

Gross Inflation (GI) is the net rate at which the general level of prices for goods and services is rising and, consequently, the purchasing power of currency is falling. Central banks attempt to limit inflation and avoid deflation to keep the economy running smoothly (World Bank WDI, 2017).

1.5.12 Deposit Interest Rate (DIR)

The interest rate paid by financial institutions to deposit account holders. Deposit accounts include certificates of deposit, savings accounts and self-directed deposit retirement accounts (World Bank WDI, 2017).

1.5.13 Gross Foreign Investment (GFI)

Gross Foreign Investment is an investment made by the companies or entities based in one country, into the companies or entities based in another country (World Bank WDI, 2017).

1.6 Objectives of the Study

The objectives of this study are as follows:

- To search the micro and macro variables which have the most rational impact on the portfolios of DSE prices;
- To propose some univariate and multivariate models for future prediction;
- To provide model adequacy and stability;
- To show the forecasting performance of the proposed model; and
- To carry out suggestions and policy recommendations.

Specific Objectives

- To propose a univariate model (s) from a various family of models—linear, and nonlinear like ARIMA, ARCH, GARCH, EGARCH, ANN, and SVM models; and
- To propose a multivariate model (s) such as Vector Auto Regression (VAR) models.

1.7 Justification of the Study

In Bangladesh, DSE plays a crucial role in industrialization and economic growth. Capital gain, efficient resource allocation, and economic growth all depend on the growth of the capital market. A few studies are conducted on the financial credibility crisis in 1996 and 2010 in the Bangladeshi stock market. This study deals with the selective macro and micro economic indicators of DSE. IMC, a micro economic indicator, and the number of TEC have a direct and immediate impact on the STR of DSE. GDP, GNI, GS, GI, DIR, the macroeconomic indicators have the indirect and long-run impact on DSE portfolios. To investigate the direct impact of DSE's turnover ratio on IMC and the number of TEC, the Cobb-Douglas (CD) production function is applied. To investigate the indirect and long-term impact, multiple linear regression models and VAR models are also applied to the indicators of DSE. To analyze the further impact on DSE, ARIMA, GARCH family models, ANN models, SVM models, and VAR models are incorporated to enrich the study.

1.8 Thesis Outline

This study is organized into five chapters. This section covers the overview of the dissertation.

Chapter 1: This chapter consists of Introduction, Statement of the Problem, Bangladesh Securities and Exchange Commission (BSEC), Dhaka Stock Exchange (DSE), Selected Indicators of DSE, Objectives of the Study, Justification of the Study, and Thesis Outline.

Chapter 2: Review of literature is discussed in this chapter. It includes the analysis of relevant literature to this study. It reviews the recent developments of stock indicators modeling and forecasting. It also concludes with the research problem that addresses the research gaps. It has made the foundation of building an efficient forecasting system of selective DSE indicators.

Chapter 3: Research methodology is assimilated in this chapter. It introduces the appropriate methodology that includes the process of modeling and forecasting strategies. Moreover, it presents a selection of indicators, univariate and multivariate models, ANN and SVM models, residuals normality and outlier testing, and finally the used software.

Chapter 4: This chapter demonstrates the results and discussion of the study. It has proposed a modeling and forecasting system of the indicators of DSE by using the supportive methods. It has analyzed the data and described the results. The key models are estimated and discussed. This chapter also includes a comparison between the models of this study and the models of other relevant studies.

Chapter 5: This chapter concludes with the results of the previous chapters. The answers to the research questions are given based on modeling and forecasting results. Furthermore, suggestions, policy recommendations, and the scope of further research are discussed.

Chapter Two

Literature Review

2. Background

This division is an attempt to find relevant literature related to the study. This is important because the present study begins at the point where the previous study ended. The chapter aims to provide a summary of the literature related to the stock market models and forecasts of other stock exchanges and DSE Limited.

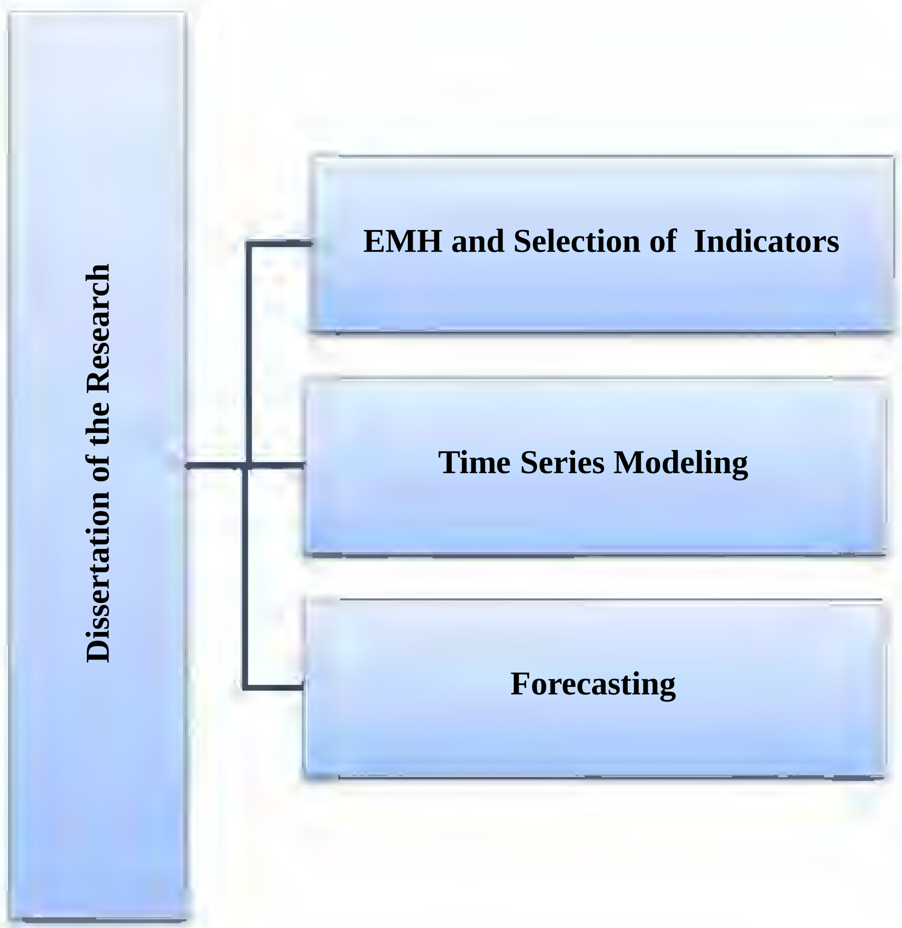


Figure 2.1 Central research theory and related areas

Stock market modeling is a theoretically significant research area to evade the monetary turmoil from the global financial crisis. The efficient market hypothesis (EMH), alternatively known as the efficient market theory, is a hypothesis that states that share prices reflect all information and consistent excess return is impossible. i.e.

- The efficient market hypothesis (EMH) or theory states that share prices reflect all available information.
- It is hypothesized that shares trade on exchanges at their fair market value.
- The proponents of EMH argue that investing in a low-cost, passive portfolio is beneficial to investors.
- Opponents of EMH believe stocks can deviate from their fair market value and that beating the market is possible.

In the areas of Efficient Market Hypothesis (EMH), selection of indicators, time series modeling, and forecasting, a reasonable number of studies are found. Figure 2.1 presents the focal theory of this dissertation as a combination of EMH and selection of indicators, time series modeling, and forecasting.

2.1 Efficient Market Hypothesis (EMH) and Selection of Indicators

According to Levine and Zervos (1998), a major stock index is considered an indicator of the economy's performance from a macro perspective. Specifically, it is shown that both stock market liquidity and bank development positively influence growth, capital accumulation, and productivity. Among the topics they studied were bank-based intermediation and market-based intermediation models. However, they did not evaluate the accuracy of the models in forecasting.

EMH modeling introduces the involvement of key indicators that play a very dynamic role in forecasting. Weak form efficiency suggests that past price, volume, and earnings data do not affect a stock's price and cannot be used to predict its future direction. i.e.

- An impossibility of predicting future prices is expressed by weak form efficiency, which is based on past values and trends.
- Efficient markets are characterized by weak form efficiency.
- According to weak form efficiency, stock prices reflect all current information.
- Technical analysis and financial advisors are little useful to advocates of weak form efficiency.

Past studies revealed that the findings of EMH on testing weak-form efficiency for developing and less developed stock markets like DSE were different. A group of researchers like Branes (1986), Chan et al. (1992), Dickinson and Muragu (1994), and Ojah and Karemera (1999) studied the weak form efficiency of the Kuala Lumpur Stock Exchange, major Asian markets, Nairobi Stock Exchange, and the four Latin American countries stock markets, respectively. Another group of researchers like Cheung et al. (1993) argued that developing and less developed stock markets are not efficient in a weak form. They applied the data from the stock market of Taiwan and Korea. Claessens et al. (1995) recommended that share prices in developing markets violate weak-form EMH. Harvey (1994) found the same findings. Roux and Gilbertson (1978) and Poshakwale (1996) exposed the properties of non-randomness in share price and the market inefficiency on the Johannesburg and Indian stock markets.

2.1.1 Efficient Market Hypothesis (EMH) of DSE

A review of all major past studies on DSE and their key findings are discussed below:

Hassan et al. (1999) analyzed the risk-return connection of the Bangladeshi stock market by using univariate daily share prices. They found positive skewness, excess kurtosis, and a lack of normality of the DSE share price. They also noticed significant serial correlation and established that the stock market is inefficient.

Mobarek and Keasey (2000) resolved that DSE does not follow a random walk model. They also found significant autocorrelation of DGI indices that is efficient in weak form. The outcomes did not vary in the case of different sub-sample of observations and excluding outliers.

Haque et al. (2001) studied the total irregular benefit from the stock market. By applying the Capital Asset Pricing Model (CAPM) and EMH, they portrayed the experience of DSE after the scam of November 1996. Based on the records of four months before and after the automation, they tested EMH. The test results demonstrated that the market did not improve, even after manipulation was continued.

Kader and Rahman (2005) did not find strong evidence of weak form efficiency of DSE by analyzing abnormal trading data by using the K% filter rule.

Islam and Khaled (2005) investigated the predictability of the share price in DSE before the boom in 1996 employing heteroskedasticity robust tests. They found the reasonable performance of short-term forecasting of DSE share prices before the boom in 1996, but they did not observe it during the post-crash periods. Based on an intensive investigation, the BSEC could have taken more transparent action.

Uddin and Alam (2007) used the Ordinary Least Square (OLS) regression to find the linear relationship among the stock price, interest rate, growth of interest rate, growth of stock price. They excluded outliers and they also found a significant negative relationship between DSE stock price and growth of interest rate.

Alam et al. (2007) depicted the DSE as an efficient market by using the stock randomness of return, stock risk-return relationships, and stock liquidity. They used CAPM to find the relationship between the risk and the expected rate of stock return of an unstable stock. The relationship was inconsequential in the DSE market. They analyzed stock risk-return, market liquidity, profit of shareholders and they also found an insignificant association among these variables.

Hossain and Kamal (2010) found a unidirectional causality between share market progress and financial growth in Bangladesh. They distinguished a comparative stochastic pattern of both the factors like share market progress and financial growth.

Ali (2011) investigated the long-run equilibrium and short-run dynamics and found a causal relationship among DGI indices, Consumer Price Index (CPI), GDP, import payment and foreign remittances. A significant cointegration among the variables was revealed. Vector Error Correction (VEC) model corrected its degree of disequilibrium by 5.98 % every month.

Shen et al. (2011) noticed a non-linear fluctuation in the stock market. The stock market was influenced by internal and external factors. Thus, stock market forecasting became a challenging job for researchers.

Hossain and Nasrin (2012) exposed that the company's selective features, reputation, net asset value, and bookkeeping data were the most influencing factors on retail investors in the stock market of Bangladesh.

Roy and Ashrafuzzaman (2015) did not predict stock price properly, but they found a rare change lying between the intrinsic value estimated by models and the actual value of the stocks. They modeled with the data series including the period 2010-2011 when the largest share market scam happened in the history of DSE.

Hasan (2015) used daily return data of DSE indices such as DSI, DGI, and DSE-20 indices from 2 January 1993 to 27 January 2013, 1 January 2002 to 31 July 2013, and 1 January 2001 to 27 January 2013, respectively. According to the random walk theory, stock price changes do not have the same distribution, and they are independent of one another. Therefore, it assumes that stock prices and markets cannot be predicted by past trends or movements. There were 4823, 2903, and 3047 daily return observations in each of these DSE indices, which did not satisfy the random walk model property. Therefore, the DSE was inefficient.

The previous studies measured DSE market efficiency by concentrating on the credibility of DSE, reliable information, consequences of financial events, sustainable policies, etc. The majority of researchers argue that DSE is inefficient or a weak form of efficiency. Nevertheless, no study focused on market efficiency based on the combination of micro and macro time series indicators of DSE. Here lacking is a crucial scope of work. This study attempts to identify the lack of EMH by various time series modeling on a trial and error basis. Evaluation of recommendations will significantly guide policymakers and regulators. The outcomes may also identify the challenges of policy implications by addressing proper stakeholders. It will open the opportunity for further studies on this issue. The following sections introduce the developments of related research.

2.1.2 Selection of Indicators of DSE

An important part of this study is the selection of stock market indicators. The most reasonable indicators of DSE portfolios are selected using the existing literature of stock market modeling and forecasting in the world. Bangladesh is an emerging developing country in South Asia. DSE is one of the major capital markets. It mobilizes savings into investments for producing goods and services, generating employment, and sustainable economic growth of the country as a whole. But any sort of disorder of the microeconomic and macroeconomic indicators may negatively affect stock prices of the capital market. A considerable amount of research literature is found demonstrating the stock market volatility through stock indicator modeling and forecasting.

Ahmed (2000) examined the significance of dividend and retained earnings to clarify the variation of the stock price in Bangladesh. The results disclosed the influence of dividend and retained earnings on the stock price. It also found the effect of typical expectation of stronger dividend and retained earnings on non-growth industries and growth industries, respectively. Thus, the dividend hypothesis was stronger than the retained earnings hypothesis.

Pu Shen (2000) investigated the association between price-earnings (P/E) ratios and the performance of consequent share markets. He found historical indications of high P/E ratios, followed by disappointing stock markets. Mainly, high P/E ratios slowed the long-run growth in stock prices. Furthermore, high price P/E ratios made the stock market profitable in the short term, but the small stockholders suffered in the long run as well.

Nelson (1976) examined the relationship between monthly share returns and inflation from 1953 to 1974 using US data. There was a negative correlation between share returns, and predicted and unpredicted inflation.

Ray (2012) found a positive association between foreign exchange reserves and share prices on the Bombay Stock Exchange (BSE), India, and the Shenzhen Stock Exchange, China.

Afzal and Hossain (2011) inspected the association between share prices and macroeconomic variables of Bangladesh. They used monthly data from July 2003 to October 2011 to check the connection between DGI, and M1, M2, inflation, and exchange rate. The result showed a long-run equilibrium relationship among the variables. It also exposed bivariate causality among the variables. Unfortunately, they failed to predict DGI, M1, M2, inflation, and exchange rates, etc.

Banerjee and Adhikary (2009) explored the dynamic association between the exchange rate from Taka to US Dollar and deposit interest rate. A notable cointegration method was used with monthly data from January 1983 to December 2006. A significant positive association was found between share prices and deposit interest rates. However, the study found a significant negative relationship between share prices and exchange rates. Furthermore, it also noticed a long-run equilibrium and causal relationship among interest rate, exchange rate, and stock return. Nevertheless, they did find any short-term effects of stock return on the interest rate and exchange rate.

A stock market index is a measurement value of stock market portfolios that is calculated from the values of registered shares. Stockholders and financial managers used it to estimate stock prices. An index is a technical term that may not be calculated directly. To develop specialized investments, financial institutions and mutual fund experts track the stock index. It also eases the share business of a nation and reflects stockholders' sentiment on the concerned economy. Thus, the present study incorporates the key indices of DSE Limited like DGI, DSEX, DSES, and DSE30.

The present study is an endeavor to investigate the econometric consequences on DSE's portfolios for the selective microeconomic and macroeconomic indicators. There are some microeconomic indicators viz. IMC and the number of TEC that influence the stock prices directly and some macroeconomic factors like GDP, GNI, GS, GI, DIR, and GFI that also influence the share prices indirectly in the long run. The activities of capital markets and the relationships between stock indices, and micro and macro-economic factors have significant importance to expose the monetary risks that make the capital market proficient.

2.2 Time Series Modeling

Data is collected over time in numerous areas of study. The sequence of observations creates a time series, for example, the closing shares prices, unemployment rate, inventory levels of production, etc. These are examples of time series data. It is used to understand the dynamics of a system that makes a sensible forecast of the forthcoming behavior of data. Therefore, many analysts and researchers are interested in modeling and forecasting time series data of the share market.

Yule (1927) studied the dynamic movement of a pendulum that is the motivation to theorize an autoregressive model for time dependency of observed values. Most physical objects show inertia and these do not change quickly over time. Thus, it counts sampling frequencies and often the successive observations are correlated over time. This correlation between serial observations is called autocorrelation.

Bizzaard and Kulahci (2011) found that most time series models based on the hypothesis of independent observations may be misguided or misused when the data are auto-correlated.

Shah et al. (2019) stated that modeling of share markets is mainly a random walk. They specified it as a fool's game to predict share prices. Forecasting share prices is a tough job because of involving an unknown number of variables. The market performs as a voting machine in the short term, but it works as a weighing machine in the long term. Thus, it makes room for modeling and forecasting share prices for a long time.

In the study and forecasting of share prices, machine learning techniques show great promise. Several relevant time series modeling pieces of literature are presented in the following subsections.

2.2.1 Linear and Nonlinear Regression Model

We need to study an alternative method that includes serial dependence among observations. In this case, linear time series models like the ARIMA model can be applied. In the case of short and medium-term forecasting, time series methods have been widely used (Box and Jenkins, 1976). ARIMA and nonlinear GARCH family models are applied for modeling and forecasting stock prices that make reasonable outcomes in the stock market.

2.2.1.1 ARIMA Modeling

The ARMA model is a vital technique for studying time series data. Yule, Slutsky, Walker, and Yaglom developed AR and MA models. ARIMA model is developed from the ARMA model by transforming non-stationary data into stationary data through differencing. It is generally used to forecast linear time series data (Chen et al., 2014).

Box and Jenkins made the ARIMA model popular among the researchers and it is known as the Box and Jenkins model (Box and Jenkins, 1976). Box and Tiao used a general transfer function engaged by the ARIMA model (Box and Tiao, 1975).

ARIMA model is referred to as the ARIMAX model when it contains other time series as input variables. Pankratz (1991) remarked on the ARIMAX as a dynamic regression model. The ARIMA technique suggests great flexibility in univariate time series models introducing the key processes like identification of models, estimation of parameters, and finally forecasting.

Hossain et al. (2015) investigated the daily volatility of the DSE market by developing a univariate ARIMA model, which presented a minimal RMSE compared to non-linear models individually. However, the mixture of ARIMA and GARCH family models had the lowest RMSE value.

Using the ARIMA model, Ayodele et al. (2014) accurately predicted the share price. Data was collected from the Nigeria Stock Exchange and the New York Stock Exchange. The ARIMA model worked well for short-term forecasting, but not for long-term forecasting.

2.2.1.2 GARCH Family Modeling

Engle (1982) developed an oversized model such as the ARCH model that handled the conditional variance of monetary time series. It is used to grab risk management, pricing derivatives, and hedging portfolios in the stock market.

Corhay and Rad (1994) used ARCH models to forecast stock prices on European capital markets. It was a satisfactory forecasting performance. However, they did not apply a mixture of ARIMA-GARCH models.

Bollerslev (1986) extended the ARCH model to the Generalized ARCH model (GARCH). Hansen and Lund (2001) also applied the extension of the GARCH model that led to the development of the EGARCH model.

Ajasi et al. (2008) connected Ghana's share markets to the exchange market using the EGARCH model. In that study, they did not evaluate out-of-sample forecasts.

Reyes (2001) observed volatility transferring among size-based share indices of the Tokyo Stock Exchange. He conducted a bivariate EGARCH model to check volatility between small and large capitalization of share indices. As well, he failed to show out sampling forecasting.

Basel et al. (2005) predicted the volatility of the S&P-500 index by using GARCH models. The out-of-sample forecasting appropriately grabs asymmetric components.

Kang et al. (2009) applied a fractionally integrated GARCH (FIGARCH) model to explore an abrupt change that reduced the long memory property of the Japanese and Korean share markets. They observed the unexpected change that was related to political and economic events. This study focused on data regarding abrupt changes in variance, which could improve predicting volatility's accuracy.

Joshi (2010) examined the volatility of closing share prices of the rising share markets of China and India. They detected the existence of non-linearity by applying the ARCH-LM test.

Ahmed and Suliman (2011) estimated the volatility of daily capital returns of the Sudan securities market using GARCH models. Similarly, Sattayatham et al. (2012) forecasted volatility and stock return of SET index in Thailand stock market using the GARCH model. However, all of them were unable to predict out-of-sample data.

Bala and Asemota (2013) dealt with month-wise data of exchange rate from US dollar, Euro, British pounds to Naira to check the volatility of exchange rate using GARCH models. Similarly, Aziz and Uddin (2014) applied GARCH models to evaluate the existence of volatility in DSE.

Bhardwaj et al. (2014) used non-structural time series models like Box-Jenkins ARIMA and GARCH models to predict the share prices of Delhi stock market. The out-of-sample forecasts were not evaluated.

Onwukwe et al. (2014) forecasted volatility of fifteen Nigerian banks using symmetrical models such as ARCH(1), ARCH(2), and GARCH(1,1) and non-symmetric models such as TARCH(1,1) and EGARCH(1,1) model. However, ARIMA models were not taken into consideration.

By applying GARCH models using high-frequency data, Hou and Li (2015) investigated CSI 300 of the Shanghai stock market, China. The results showed a one-way transmission of the volatility of the CSI 300 index. Moreover, the future index of CSI 300 indicated an efficient spot market in terms of information.

Sharma and Vipul (2015) applied the GARCH model with diversified samples, asset types, and performance assessment criteria for forecasting financial indices of stock markets. These results accurately forecasted the NIFTY index of India using GARCH family models. It also brought attention to selecting the accurate benchmark and loss criteria. However, the ARIMA model was not considered.

Hou and Li (2015) argued that the GARCH model, a heavily parameterized model, could capture and forecast the numerous dimensions of volatility in the stock market.

In step with Ali and Mhmoud (2013), and Vee and Gonpot (2011) forecasted volatility of exchange rate from Mauritian Rupee to US Dollar. They applied the GARCH(1,1) model considering Generalized Error Distribution (GED) and Student's-t distribution.

Ahmed and Shabri (2013) fitted GARCH family models for predicting the prices of crude. GED, student's t, and normal error distribution were assumed to estimate the parameters of GARCH family models. Moreover, plenty of experiential studies are done applying GARCH family models considering normal error distribution.

2.2.1.3 VAR Modeling

Vector autoregression (VAR) model, an extension of the univariate AR process, is an optimistic model to analyze the multivariate time series data. Forecasting of financial time series has become very popular among researchers. VAR is a suitable model for unfolding the dynamic characteristics of a monetary time series. Simultaneous equations are used to estimate the parameters of VAR models (Sims, 1980). In the case of the VAR model, forecasting is relatively flexible as it makes restrictions on future paths of specific variables. VAR models with multiple variables like stocks, bonds, and foreign exchange rates examined the association between risk and return (Culbertson, 1996). However, a few studies are conducted about the credibility crisis of Bangladeshi financial market, especially in 1996 and from 2010 to 2013. Moreover, the VAR model draws the essential inference and makes policies. Certain assumptions need to be imposed to investigate the causal arrangement of the data. The causal influences of unexpected variables on the dependent variables are summarized by impulse response functions and forecast error variance decompositions (Lutkepohl, 1991; Watson, 1994; Hamilton, 1994; Campbell, 1997; Waggoner and Zha, 1999; Lutkepohl, 1999; Mills, 1999; Tsay, 2001).

Hossain et al. (2015) developed VAR models for predicting selected micro economic indicators such as stock market index, stock trade, invested market capital, and stock volume of the DSE for a long-term period. They have found the minimum RMSE.

2.2.1.4 ANN Modeling

Neural Network (NN) is an artificial intelligence scheme (Kai and Wenhua, 1997). NN is a time-based, large-scale, and non-linear dynamic system. NN connects many nodes of a weight matrix. This technique forecasts share prices of the stock market properly. Backpropagation neural networks include three layers such as input, hidden, and output layers. To capture uncertain and non-robust data is the vital advantage of NN models. Because of this, NN is popular for predicting share indices and analyzing stock prices. NN is also effective for the unavailability of certain data. The combinations of numerous NN forecast enormous value of a time series properly, as the combined methods emphasized different features of data set that are essential for computing output (Afolabi and Olude, 2007).

Backpropagation neural networks and ARIMA models were used to predict trading volume, and it was observed that NN predicting capacity was rationally well relative to the ARIMA model (Kaastra and Boyd, 1995). Furthermore, the artificial NN models forecast daily exchange rates accurately (Hans and Kasper, 1998).

2.2.1.5 SVM Modeling

Vapnik (1979) developed the Support Vector Machine (SVM). Burges (1998) argued that SVM is a suitable method for data classification. Smola and Schölkopf (2004) applied SVM as a regression model. Müller et al. (1997) and Kim (2003) used the SVM regression model for forecasting purposes.

In several situations, like classification, pattern recognition, and regression analysis, the SVM model outstandingly performs well based on the error rates on test samples (Burges, 1998). The acceptance and extraordinary performance of the SVM model are described by proper formulation of a convex objective function with constraints by Lagrange multipliers, decision function, and kernel functions (Vapnik, 1998; Vapnik, 1999).

Moreover, in the background of time series modeling on financial data, the SVM model is applied for the subsequent justifications: (a) data could be performed without prior assumptions; (b) conventional NN models are applied using empirical risk minimization principle, whereas SVM model works on structural risk minimization principle that shows resilient to the over-fitting problem; (c) SVM model linearly constrained quadratic program so that it can make objective function optimum, while NN models may fail to do so (Kim, 2003; Huang et al., 2005).

Though the SVM model shows outstanding performance and encouraging results to forecast financial data, a few studies were conducted on forecasting financial time series by applying the SVM regression-based model compared with others models. Moreover, numerous studies revealed that SVM performed better than time series techniques, Backpropagation NN, and ARIMA models (Tay and Cao, 2001; Kim, 2003; Thissen et al., 2003).

Chapter Three

Methodology

3.1 Introduction

The use of time series modeling and forecasting in stock market analysis has become very popular. Time series research has become an integral part of many disciplines, from economics to industry to engineering and finance to politics and ecology. The main aim of time series modeling is to search for the most reasonable variables and models based on a trial and error process. This chapter employs the selection of reasonable variables and proper models. In particular, modeling and forecasting are described. The software used, as well as data adjustment, is also discussed.

3.2 Source of Data

This study is conducted based on the secondary time series data. Secondary time series data like annual IMC (US$), number of TEC and STR during 1990 to 2012, and GDP, GNI, GS, GI, DIR, and GFI during 2005 to 2012 were collected from the World Bank website (Source: http://data.worldbank.org/country/bangladesh). Daily time series data like Capital, Volume, Value, Trade and DGI (All share prices of A, B, G & N categories of the portfolios of DSE) during June 2004 to July 2013, and DSEX, DSES, and DSE30 Indices during January 2014 to December 2018 were collected from DSE website (Source: http://www.dsebd.org/recent_market_information.php). To check the seasonality and month-wise forecasting performance, Capital, Volume, Value, Trade, DGI, DSEX, DSES, and DSE30 Indices were transformed to a monthly scale.

3.3 Short Term and Long Term Analysis of DSE Indicators

To analyze the short-term impacts on DSE portfolios, the micro economic variables are considered to model the relevant time series data, and to analyze the long-term impacts on DSE portfolios the macroeconomic variables are considered to model the relevant time series data. This study explores the effect of Dhaka

Stock Exchange (DSE) portfolios corresponding to Bangladesh's selective macro and micro economic indicators. The microeconomic indicators like trade, market volume, invested capital, and market value have a direct impact on DSE portfolios and the macroeconomic indicators are GDP, GNI, GS, GFI, and GI have an indirect and long-run impact on DSE portfolios. The data flow diagram of micro and macro indicators of DSE portfolios on DGI is shown in Figure 3.1.

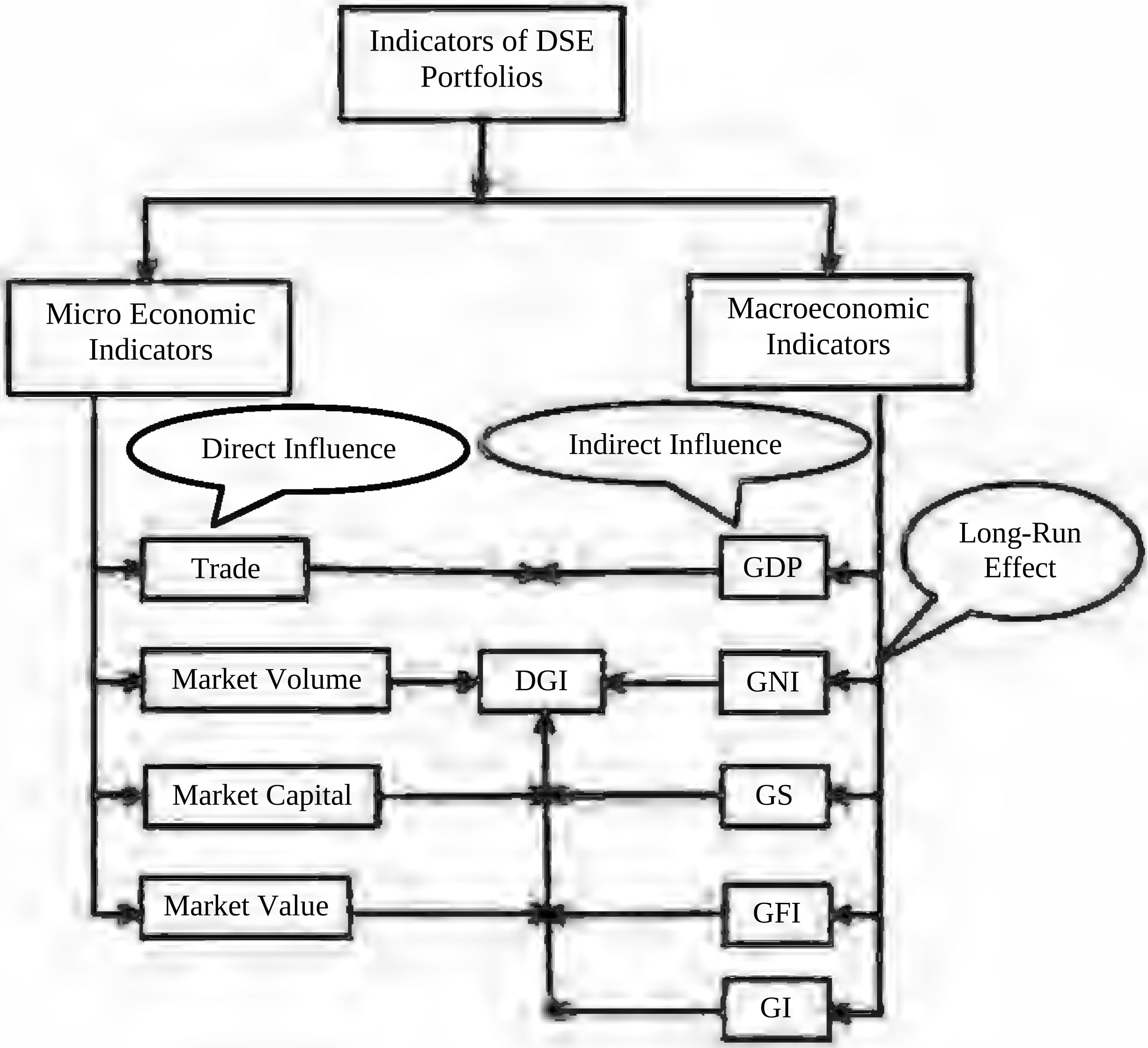


Figure 3.1 Indicators of the DSE's portfolios

3.4 Modeling and Forecasting Strategy

One of the most difficult and critical jobs among researchers is efficient financial time series forecasting. Some models are useful for reasonable forecasting of a certain data series from different classes of models. So, it needs to give priority to fit a proper model based on a trial and error process. Firstly, time series data are explored using visual inspection like time series plot and then trend and seasonality are identified. To make proper univariate ARIMA models, stationary properties are ensured and non-stationary series are transformed to make stationary. ARIMA models, GARCH family models, ANN models, and SVM models are estimated with approximately 75% data as the training samples and with 25% data as the test samples. To find the reasonable ARIMA and GARCH family models, the minimum value of AIC, BIC, and maximum value of R-square is considered. Finally, forecasting is considered with the proposed models which produce a minimum Mean Square Error (MSE) than that of other models. The functional diagram of the modeling and forecasting strategy is presented in Figure 3.2. The entire modeling and forecasting process is classified into three steps—i) Identifications of models, ii) Estimations of models, and iii) Forecast using the proposed models. The details of modeling and forecasting are discussed in the following section.

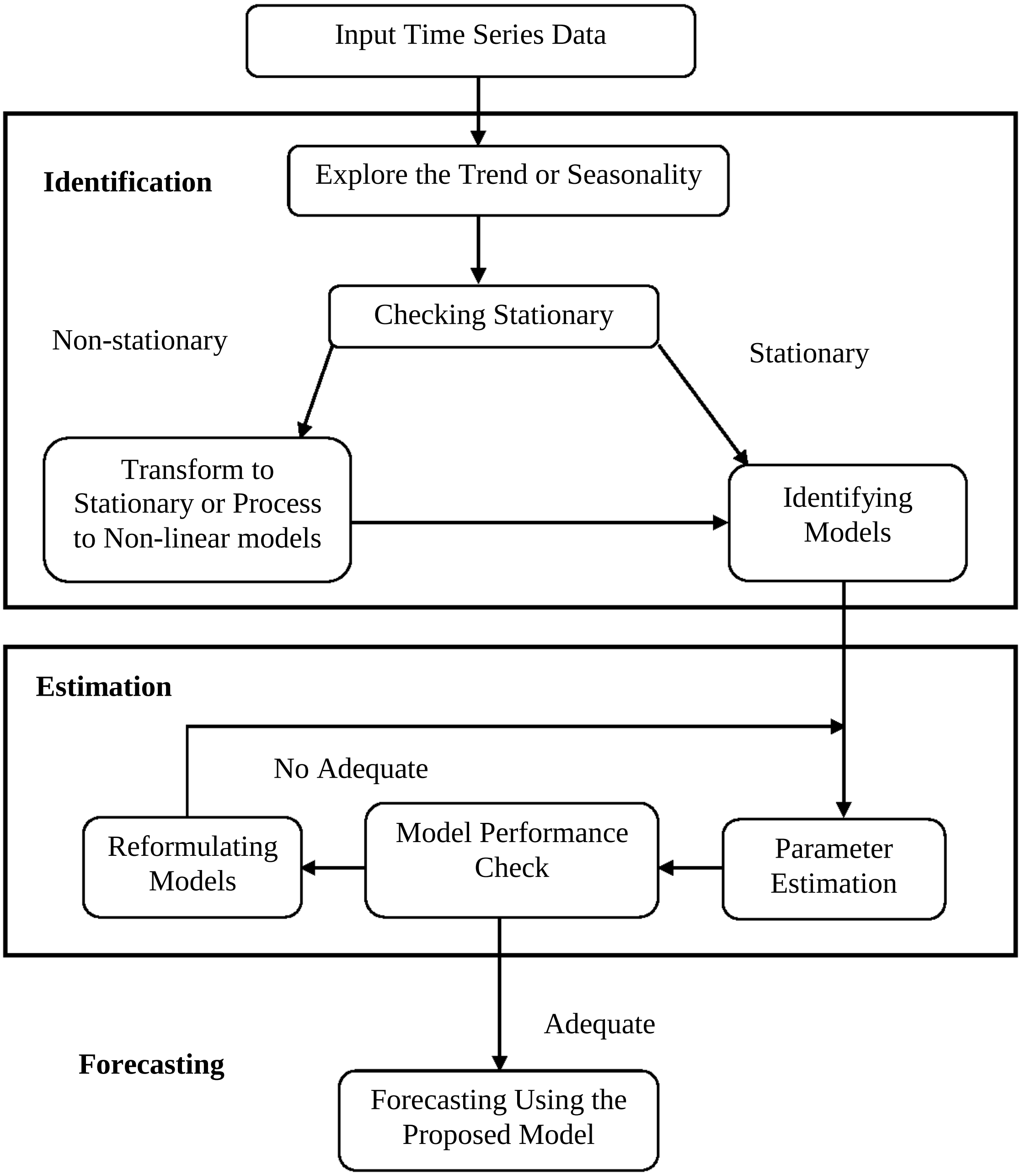


Figure 3.2: Functional diagrams of modeling and forecasting strategy

3.4.1 Stationary Process

A system called the stationary process is based on the evidence that the method is in a specific state of statistical equilibrium. Stationary models are considered an integral class of stochastic models for the description of time series, which has received a lot of attention. It assumes that the process remains in equilibrium at a constant mean level. Two types of stationary are available one is in the weak sense and another is in the strict sense. Generally, if the mean and variance of a stochastic process are constant over time and covariance of two time periods relies only on the lag between two time periods and not on the actual time at which the covariance is measured. Then, the process is called stationary. Suppose Y_t be a stochastic time series. Then Y_t is said to be a stationary process if it satisfies the following properties:

- The first order moment (mean) exists i.e., $E(Y_t) = \mu \ \forall t$
- The variance is constant through time i.e., $Var(Y_t) = E(Y_t-\mu) = \sigma^2 \ \forall t$
- Covariance, $\gamma_k = Cov(Y_t, Y_{t+k}) = E[(Y_t-\mu)(Y_{t+k}-\mu)]$ does not depend on the time t.

The above definition of stationary is based on a weak sense. If its properties are unchanged by a shift in time origin, a stochastic process is said to be purely stationary; that is if the joint probability distribution associated with n observations (Y_{t1}, Y_{t2}, ..., Y_{tn}), made at time t_1, t_2, ..., t_n is the same as that associated with n observations (Y_{t1}, Y_{t2}, . . ., Y_{tn}), made at time t_{1+k}, t_{2+k}, ..., t_{n+k}. Thus, a discrete approach is stationary; the joint distribution of any set of observations must not be changed by shifting forward or backward times of observations at any integer amount k.

3.4.2 Multiple Regression Model

The Cobb-Douglas (CD) functional form of production functions is commonly used in economics to describe the relation of output to inputs. It was proposed by Knut Wicksell (1851 - 1926), and it was evaluated in 1928 by Charles Cobb and Paul Douglas against observational evidence. They calculated a simple view of the economy in which the amount of labor involved and the amount of capital invested was used as the independent variables for estimation of production output. This model is exceptionally reliable, as several other variables that are impacting economic growth. This nonlinear function is used to model for stock traded turnover ratio prediction that is of the following form:

$$Y = CL^{\alpha} K^{\beta} \tag{3.1}$$

where:

Y = Total stock traded turnover ratio

L = Total number of the enlisted companies in the stock market

K = Total invested capital of the stock market

C = Total factor turnover ratio of the stock market

and, α and β are the output elasticity of the total number of enlisted companies and total invested capital of the stock market, respectively. These values are constants. Taking log on both sides of equation (3.1), it becomes a simple log multiple linear regression of the following form:

$$\text{Log}Y = c + \alpha \text{Log}L + \beta \text{Log}K + \varepsilon \tag{3.2}$$

where:

ε = Random error term

The hypothesis of constant returns to scale is then tested by the restriction:

$\alpha+\beta = 1$.

3.4.3 ARIMA Model

The concept of Autoregressive Integrated Moving Average (ARIMA) is nothing but a mixture of three systems, the process of Autoregressive (AR), the process of Moving Average (MA), and the integrated process. ARIMA (p, d, q) denotes the general ARIMA order of p, d, and q, and it can be written concisely:

$$\phi(B)\nabla^{d}Y_t = C + \theta(B)\varepsilon_t \tag{3.3}$$

$\nabla^{d} = (1-\text{d})^{\text{d}}$ (The d order differencing operator)

$\phi(B) = (1-\phi_1 B - \phi_2 B^2 - \phi_3 B^3 - \cdots - \phi_p B^p)$ (The p order AR operator)

$\theta(B) = (1-\theta_1 B - \theta_2 B^2 - \theta_3 B^3 - \cdots - \theta_p B^p)$ (The q order MA operator)

ε_t = Random Shocks, C is the constant and Y_t is any time series.

To achieve stationary, the difference is not necessary, d = 0 and the model is simplified to ARMA.

3.4.4 GARCH Family Models

There are a lot of GARCH models which are used for financial time series forecasting. In this study, the following models are estimated and analyzed.

3.4.4.1 ARCH Model

Extensive work was dedicated over the past three decades for modeling and forecasting the equity returns as well as other financial time series. The typical Autoregressive Conditional Heteroskedasticity (ARCH) model was introduced by Engle (1982). The ARCH model of order q, denoted by ARCH(q), can be defined based on past innovations of conditional variance as follows:

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2 \qquad (3.4)$$

where, ε_t denotes a discrete-time stochastic taking the form of $\varepsilon_t = z\sigma_t$ and z_t ~ iid (0,1), and σ_t is the normal conditional standard deviation of return at time t.

3.4.4.2 GARCH Model

Another extension suggested by Bollerslev (1986), known as the Generalized ARCH (GARCH) model, implies that both past disruptions and past volatility are a feature of the time-varying volatility model. The model GARCH(p, q) is an ARCH model of an infinite order expressed by:

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{p} \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^2 \qquad (3.5)$$

where, α_0, α, and β are non-negative constants. In order to describe the GARCH process, it is required that $\alpha > 0$.

3.4.4.3 EGARCH Model

Nelson (1991) introduced the EGARCH or Exponential GARCH model. For the conditional variance, the specification is:

$$\log \sigma_t^2 = \omega + \beta \log \sigma_{t-1}^2 + \alpha \left| \frac{\varepsilon_{t-1}}{\sigma_{t-1}} \right| + \gamma \frac{\varepsilon_{t-1}}{\sigma_{t-1}} \qquad (3.6).$$

Note that the conditional variance is on the left-hand side. This means that the leverage effect is proportional, not quadratic and that the conditional variance estimates are guaranteed to be non-negative. The existence of leverage effects can be checked by the principle that $\gamma > 0$. The result is asymmetrical if it is $\gamma \neq 0$. Nelson claims that the ε follows a generalized error distribution. The specification of Nelson for log conditional variances marginally varies from the above specification:

$$\log \sigma_t^2 = \omega + \beta \log \sigma_{t-1}^2 + \alpha \left| \frac{\varepsilon_{t-1}}{\sigma_{t-1}} - \sqrt{\frac{2}{\pi}} \right| + \gamma \frac{\varepsilon_{t-1}}{\sigma_{t-1}} \qquad (3.7).$$

Under the assumption of normal errors, this model will yield similar estimates to those that vary by $\alpha\sqrt{2/\pi}$, except for the intercept term, ω:

$$\log \sigma_t^2 = \omega + \sum_{i=1}^{p} \beta_i \log \sigma_{t-1}^2 + \sum_{j=1}^{q} \left(\alpha_j \left| \frac{\varepsilon_{t-j}}{\sigma_{t-j}} - \sqrt{\frac{2}{\pi}} \right| + \gamma_j \frac{\varepsilon_{t-j}}{\sigma_{t-j}} \right) \qquad (3.8).$$

The expression for the leverage effect, denoted as γ, is negative and statistically distinct from zero, suggesting the presence of the leverage effect during the sample period in future stock returns.

3.4.5 VAR Models

The time series Y_t fits a model of VAR(p) if it satisfies

$$Y_t = \Phi_0 + \Phi_1 Y_{t-1} + \cdots + \Phi_p Y_{t-p} + \alpha_t\,, \;\; p > 0 \qquad (3.9)$$

where, Y_t is a vector of the response variable, Φ_0 is a k-dimensional vector, and α_t is a sequence of serially uncorrelated random vectors with mean zero and covariance matrix Σ. Covariance matrix Σ must be positive definite; otherwise, the dimension of Y_t can be often reduced. The error term, α_t is a multivariate normal, and Φ_j are $k \times k$ matrices. By the back-shift operator *B*, the VAR(p) model can be expressed as

$$(I - \Phi_1 B - \cdots - \Phi_P B^P)Y_t = \phi_0 + \alpha_t$$

Where, *I* be the k×k identity matrix. In a compact form as follows:

$$\phi(B)Y_t = \phi_0 + \alpha_t$$

Where, $\phi(B) = I - \Phi_1 B - \cdots - \Phi_P B^P$ is a matrix polynomial.

Consider the following consecutive VAR models:

" $Y_t = \phi_0 + \Phi_1 Y_{t-1} + \alpha_t$

$Y_t = \phi_0 + \Phi_1 Y_{t-1} + \Phi_2 Y_2 + \alpha_t$

… = …

$$Y_t = \phi_0 + \Phi_1 Y_{t-1} + \cdots + \Phi_p Y_{t-p} + \alpha_t \qquad (3.10).$$

… = … ‘’

For the parameter estimation of these models, the ordinary least squares (OLS) approach is used. In multivariate statistical analysis, this is called multivariate linear regression estimation (Tsay, 2001). For the i^{th} equation in Eq. (3.10), let, $\widehat{\Phi}_j^{(i)}$ be the OLS estimate of Φ_j and $\widehat{\Phi}_j^{(i)}$ be the estimate of ϕ_0, where the superscript (*i*) is used to represent that the estimates are for a VAR(*i*) model. Then the residual is

$$\hat{\alpha}_t^{(i)} = Y_t - \widehat{\Phi}_1^{(i)} Y_{t-1} - \cdots - \widehat{\Phi}_1^{(i)} Y_{t-i}$$

For $i = 0$, the residual is defined as $\hat{Y}_t^{(0)} = Y_t - \bar{Y}$, where $\bar{Y}$ is the sample mean of Y_t. The residual covariance matrix is defined as

$$\hat{\Sigma}_i = \frac{1}{T-2i-1} \sum_{t=i+1}^{T} \hat{\alpha}_t^{(i)} (\hat{\alpha}_t^{(i)})^{\intercal} \qquad (3.11).$$

AIC is used to select the order of the VAR model.

3.4.5.1 Structural Analysis by Impulse Response Functions

The common structure of the VAR(p) model has been expressed as a representation of Wold as follows:

$$Y_t = \mu + \theta_i \alpha_{i-1} + \theta_2 \alpha_{t-1} + \cdots \qquad (3.12)$$

where, θ_s are the *nxn* matrices. To interpret the (i,j)-th element θ_{ij}^{s}, element of the matrix θ_s as the dynamic multiplier or impulse response

$$\frac{\delta y_{i,t+s}}{\delta \alpha_{j,t}} = \frac{\delta y_{i,t}}{\delta \alpha_{j,t-s}} = \theta_{ij}^{s} \quad i,j = 1,2,\ldots,n \qquad (3.13).$$

The criteria for the eq. (3.9) is a diagonal matrix Σ = var (α_t). If Σ is diagonal, it reveals the uncorrelated components of Σ and α_t. One way to render the uncorrelated errors is to calculate the triangular structure of the VAR(p) model.

$$'' y_{1t} = c_1 + \acute{\alpha}_{11} Y_{t-1} + \cdots + \acute{\alpha}_{1p} Y_{t-p} + \eta_{1t}$$

$$y_{2t} = c_1 + \beta_{21} Y_{1t} + \acute{\alpha}_{21} Y_{t-1} \ldots + \acute{\alpha}_{2p} Y_{t-p} + \eta_{2t}$$

$$\ldots = \ldots$$

$$y_{nt} = c_1 + \beta_{n1} Y_{1t} + \beta_{n,n-1} Y_{n-1,t} + \acute{\alpha}_{n1} Y_{t-1} \ldots + \acute{\alpha}_{np} Y_{t-p} + \eta_{nt} '' \qquad (3.14).$$

The calculated covariance matrix is diagonal for the error vector η_t. The uncorrelated errors η_t are known as structural errors. The Wold representation of Y_t is based on the following orthogonal errors:

$$Y_t = \mu + \Theta_0 \eta_t + \Theta_1 \eta_{t-1} + \Theta_2 \eta_{t-2} + \cdots \; .$$

3.4.6 ARIMA with GARCH Family Model

The finite mixture of the ARIMA-GARCH Model for time series Y_t is expressed as the following form:

$$Y_t = C + \sum_{i=1}^{I} \theta_i Y_{t-i} + \sum_{j=1}^{J} \phi_j \in_{t-j} + \sigma_t Z_t \qquad (3.15)$$

$$\sigma_t^2 = K + \sum_{q=1}^{Q} A_q \sigma_{t-q}^2 + \sum_{p=1}^{p} G_p \in_{t-p}^2 \qquad (3.16)$$

where, k>0, $A_q \geq 0$, $G_p \geq 0$ and σ_t^2 is the conditional variance, Z_t is the standardized independent and identically distributed *(i.i.d)* random variable drawn from some indicated probability distribution; Z_t follows N(0,1) with mean zero and variance unity or student t-distribution with the degree of freedom *v*. ML - ARCH (Marquardt) algorithm is used to estimate the mixture model. We estimate tail quantities by assuming normal error distribution, multiplying the estimate of σ_t

with the standard quartiles of each distribution, and lastly adding the conditional mean. Similarly for EGARCH model, σ_t^2 can be expressed as:

$$\log\sigma_t^2 = \omega + \beta\log\sigma_{t-1}^2 + \alpha\left|\frac{\varepsilon_{t-1}}{\sigma_{t-1}}\right| + \gamma\frac{\varepsilon_{t-1}}{\sigma_{t-1}} \quad (3.17).$$

The leverage effects can be tested by the hypothesis that $\gamma > 0$. The influence is asymmetric if, $\gamma \neq 0$.

3.4.7 ANNs Models

Flexible computational modeling of an extensive variety of nonlinear problems is Artificial Neural Networks (ANNs) (Wong et al., 2000). There are three layers of neurons in an ANN model: the input layer where the data reaches the system, the hidden layer where the data is interpreted, and the output layer where the system determines based on the data what to do. It is a form of machine learning that models the human brain and consists of a set of artificial neurons. Neurons have fewer similarities in ANNs than biological neurons. A number of inputs are accepted by each neuron in ANNs. These inputs use an activation function that results in the activation level of the neuron (output value of the neuron). The neuron is the basic information processing unit of ANNs models. Let Y_t is a predictable (dependent variable) time series which consists of a set of links, describing the neuron inputs, with weights $W_1, W_2, \ldots, W_i$. An adder function (linear combiner) for computing the weighted sum of the inputs $X_1, X_2, \ldots, X_i$ (independent variables) formulated as:

$$\mathrm{u} = \sum_{i=1}^{\mathrm{m}} w_i x_i \quad (3.18)$$

where, the inputs $X_1, X_2, \ldots, X_i$ represent one period, two period, …, i^{th} period past lagged time series data of Y_t. The functional structure of ANN model is shown by Figure 3.3. The activation function, $f(.)$ limiting the amplitude of the neuron output and '*b*' denotes bias as follows:

$$\mathrm{y} = f(\mathrm{u} + b) \quad (3.19).$$

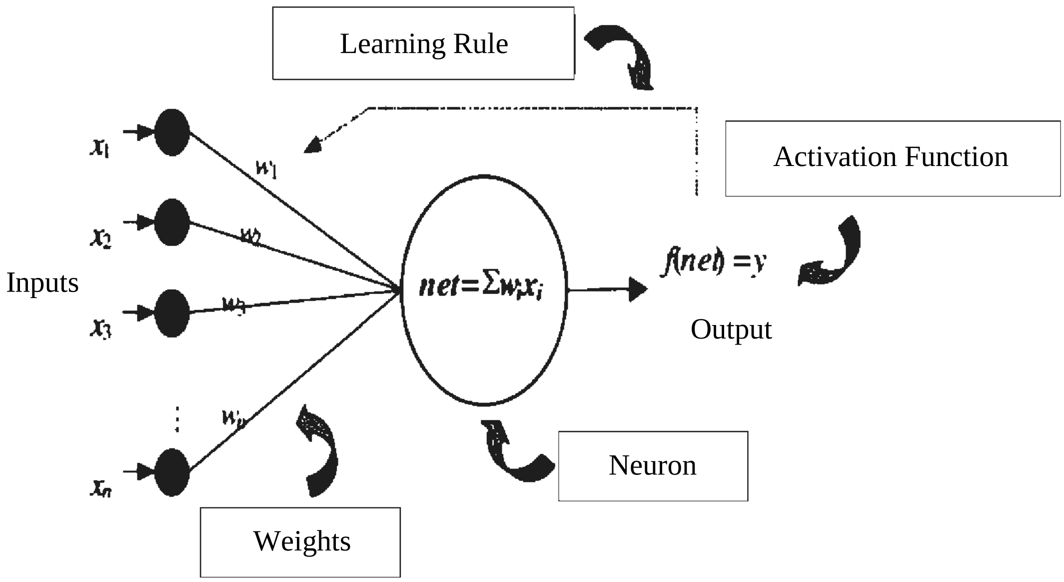


Figure 3.3: Functional diagrams of ANNs structure

The bias is called an external parameter of the neuron. It can be modeled by adding an extra input v that is called an induced field of the neuron. It can be expressed as:

$$''v = \sum_{i=0}^{m} wixi \qquad (3.20)$$
$$w_0 = b''$$

The choice of the activation function f (.) determines the neuron model based on the following functional form:

Step function: $f(v) = \begin{cases} a & \text{if } v < c \\ b & \text{if } v > c \end{cases}$,

Ramp function: $f(v) = \begin{cases} a & \text{if } v < c \\ b & \text{if } v > d \\ a + ((v-c)(b-a)/(d-c)) & \text{otherwise} \end{cases}$, Sigmoid function with z, x, y parameters: $\varphi(v) = z + \frac{1}{1+\exp(-xv+y)}$ and

Gaussian function: $f(v) = \frac{1}{\sqrt{2\pi\sigma}}\exp\left(-\frac{1}{2}\left(\frac{v-\mu}{\sigma}\right)^2\right)$.

3.4.8 SVM Models

The basic principle of Support Vector Machines (SVM) for function approximation is to transform the data x into a nonlinear mapping into a high dimensional feature space and then perform a linear regression in the feature space. Let us consider a training set of n data points $\{x_i, y_i\}_{i=1}^{n}$ with input data $x_i \in R^P$, is the total number of data patterns and output $y_i \in R$. The SVM approximate the function in the following form:

$$y(x) = w^T \emptyset(x) + b \quad (3.21)$$

where, $\emptyset(x)$ represents the high dimensional feature spaces, that is nonlinearly mapped from the input space x. By minimizing the regularized function, the coefficients w and b are estimated:

$$R(c) = \frac{1}{2}\|W\|^2 + \frac{C}{n}\sum_{i=1}^{n} L_\varepsilon(d_i, y_i) \quad (3.22)$$ and

$$L_\varepsilon(d_i, y_i) = \begin{cases} |d_i - y_i| - \varepsilon, \text{ when } |d_i - y_i| \geq \varepsilon \text{ and} \\ 0, \; otherwise \end{cases} \quad (3.23).$$

To estimate w and b, Eq. (3.22) is transformed to the primal function given by Eq. (3.24) with introducing the positive slack variables ξ and $\xi*$ as follows:

‘’ Minimize $R(w, \xi *) = \frac{1}{2}\|W\|^2 + C\sum_{i=1}^{n}(\xi_i + \xi_i *)$

$$\text{Subject to} \begin{cases} d_i - w\phi(x_i) - b_i \leq \varepsilon + \zeta_i \\ w\phi(x_i) + b_i \leq \varepsilon + \zeta_i * \\ \zeta_i, \zeta_i * \geq 0 \end{cases} \text{’’} \quad (3.24).$$

The first term $\frac{1}{2}\|W\|^2$ is referred to as the norm of the weights vector, d_i is the desired value, and C is referred to as the regularized constant that defines the tradeoff between the empirical error and the regularized term. The SVM tube size is ε and it corresponds to the estimated precision imposed on the data points in the training period. Now, the slack variables ζ and ζ_i* are included. By considering Lagrange multipliers and exploiting the optimality constraints, the decision function has the subsequent explicit form expressed by Eq. (3.21).

$$y(x) = \sum_{i=1}^{n}(a_i - a_i*)K(x, x_i) + b \qquad (3.25).$$

In Eq. (3.25), a_i and a_i* termed as Lagrange multipliers. These multipliers satisfy the equalities $a_i \times a_i* = 0$, $a_i \geq 0$ and $a_i* \geq 0$. Where, $i = 1, 2, ..., n$ is obtained by maximizing the dual function of Eq. (3.24) which has the following form:

$$''R(a_i, a_i*) = \sum_{i=1}^{n} d(a_i - a_i*) - \varepsilon\sum_{i=1}^{n}(a_i + a_i*) - \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}(a_i - a_i*)(a_j - a_j*)k(x_i, x_j)$$

$$\sum_{i=1}^{n} a_i = \sum_{i=1}^{n} a_i* \qquad (3.26).$$

with the constraints $0 \leq a_i \leq C,\ i = 1,\ 2,\ ...,\ n$ ”

$$0 \leq a_i* \leq C,\ i = 1,\ 2,\ ...,\ n$$

$K(x_i, x_j)$ is defined as the kernel function. The kernel is equal to the inner product of two vectors and x_i and x_j, in the feature space $\phi(x_i)$ and $\phi(x_j)$, that is, $K(x_i, x_j) = \phi(x_i) \times \phi(x_j)$. The typical illustrations of the kernel function are as follows:

Linear: $K(x_i, x_j) = x_i^T x_j$,

Sigmoid: $K(x_i, x_j) = \tanh(\gamma x_i^T x_j + r)$,

Polynomial: $K(x_i, x_j) = (\gamma x_i^T x_j + r)^d$, $\gamma > 0$ and

Radial basis function (RBF): $K(x_i, x_j) = \exp(-\gamma\|x_i - x_j\|^2)$, $\gamma > 0$

where, γ, r and d are kernel parameters. The parameters of the kernel should be carefully chosen as it indirectly describes the arrangement of high-dimensional feature space $\phi(x)$ and hence it controls the complexity of the final solution. The SVM architecture is shown in Figure 3.4. The inputs x_{t-1}, x_{t-2}, ..., x_{t-p} represent one period, two periods, …, p^{th} period past lagged time series data of y_t, respectively.

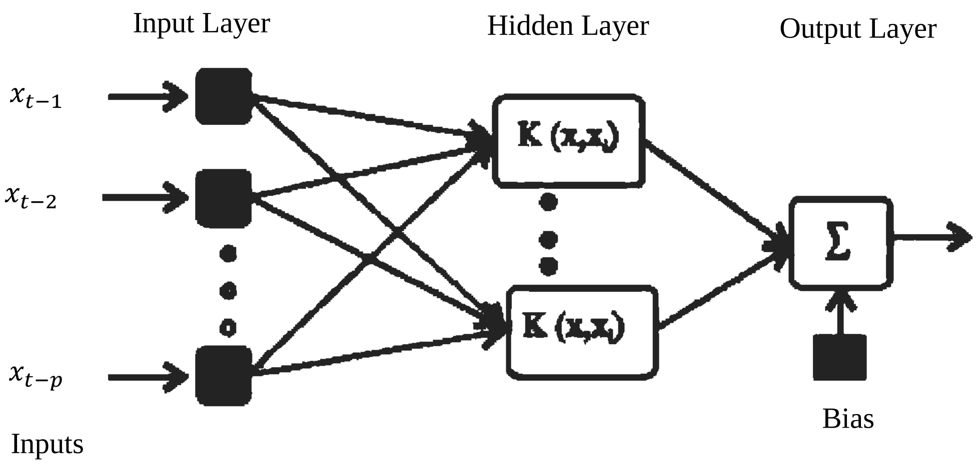


Figure 3.4: Functional diagrams of SVM architecture

3.5 Model Selection Criteria

This is a very tough job to select the best algorithm. Real data do not follow any particular model. The general instruction is that: firstly we have to select what measure of forecast error is most suitable for the particular situation at hand. Mean Squared Error (MSE) and Mean Absolute Error (MAE) are generally used for model selection. The mathematical formulas are as follows:

$$MSE = \frac{1}{n}\sum_{t=1}^{n}\left(Y_t - \hat{Y}_t\right)^2 , \quad RMSE = \sqrt{\frac{1}{n}\sum_{t=1}^{n}\left(Y_t - \hat{Y}_t\right)^2} \quad \text{and} \quad MAE = \frac{1}{n}\sum_{t=1}^{n}\left|Y_t - \hat{Y}_t\right|$$

where, Y_t = Observed value at time t and $\hat{Y}_t$ = Forecasted value at time t. We prefer RMSE to compare the performance among the models. There are some other statistics for model selection criteria like Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC) (Schwarz, 1978), and Schwarz Criterion (SC), which is closely related to AIC. We choose this model that gives the smallest value of these criteria. Also, there are some disputes among the econometricians about which criteria perform better. However, in this view, we may use all of these model selection criteria.

AIC (Akaike, 1974) is one of the most important criteria for checking the adequacy as well as the lag order of a model. AIC is defined as:

$$AIC = \log\left(\frac{\sum \hat{\varepsilon}_i^2}{N}\right) + \frac{2k}{N}$$

where, $\sum \hat{\varepsilon}_i^2$ is the sum of squared residuals. In theory, the AIC approaches a minimum value by increasing the number of lags up to the point. Thus one can choose a lag structure.

Another penalized maximum likelihood criterion is BIC. Schwarz first introduced it in 1978. In a Bayesian context, BIC was derived and approximated a version of the process from Laplace. The criterion for the BIC is minimized as

$$BIC = \log\left(\hat{\sigma}^2\right) + \frac{n\log(N)}{N}$$

where, n is the dimensionality of the model,

$(1-\phi_1 L - \phi_2 L^2 - \cdots - \phi_p L^p) y_t = c + (1-\theta_1 L - \theta_2 L^2 - \cdots - \theta_q L^q)\varepsilon_t$ is the estimate of the variance and *N* is the sample size.

3.6 Diagnostic Checking

Diagnostic checks are done in order to diagnose a potential lack of fit. The model is ready for use if no lack of fitting is demonstrated. The iterative step of identification, estimation, and diagnostic checking is replicated until a suitable model is established. The diagnostic checks are conducted in two ways—(i) The pre-test checks that are conducted before model estimating; (ii) The post-test checks that are conducted after model estimating. Stationary tests are conducted in the pre-test stage, and autocorrelation, normality, and outlier of the residuals are conducted in the post-test stage. The statistical methods and tests are conducted for possible diagnostic checking in the following subsection.

3.6.1 Differencing Method

The process of differencing a time series consists of subtracting the values of the measurements from each other in some arranged time-dependent order. Differencing is a successive change in the series for all values. For instance, a transformation of the first-order difference is defined as the difference between the values of two adjacent observations. The second-order difference consists of the differencing of the first order differencing series and so on.

There are two types of differencing—

(i) Non-seasonal differencing
(ii) Seasonal differencing

(i) Non-seasonal differencing: The change of time series (Y_t) between the values that are parted by just one time period is called non-seasonal differencing or regular differencing. For instance, the transformation of the first order differencing is defined as the difference between the values of two adjacent observations; the second order differencing is the difference of the differencing series, etc. We calculate the successive differences between observations parted by one time period in order to perform non-seasonal differencing.

Let, Y_t = original series,

D_t = first order difference series, and

D_{2t} = second order difference series.

Hence, $D_t = Y_t - Y_{t-1}$

and $D_{2t} = D_t - D_{t-1} = (Y_t - Y_{t-1}) - (Y_t - Y_{t-2}) = Y_t - 2Y_{t-1} + Y_{t-2}$ and so on.

(ii) Seasonal differencing: A series changes in a seasonal manner that is referred to as seasonal differencing. It generally induces a constant mean. We measure the successive changes between observations separated by s time periods to perform seasonal differencing, s is the number of seasons, for quarterly data $s = 4$, for monthly data $s = 12$, and so on. A series can be differenced non-seasonally only, seasonally only, or both ways. Let D denote the order of seasonal differencing. If d = 0, a seasonal differencing series ($D = 1$) is calculated for all t as:

$D_t = Y_t - Y_{t-s}$

Nearly always, setting $D = 1$ eliminates any large seasonal shifts in the level of the series. If both non-seasonal and seasonal differencing are used, either one may be done first; the result is always identical.

3.6.2 Autocorrelation

In a single time series Y_t, autocorrelation measures the direction (positive or negative) and intensity of the relationship among observations when the observations are split by k time breaks, for $k = 1, 2, 3, 4, ..., K$. To construct the column Y_{t+k}, we offset the column of Y_t observations by k time breaks for every k. We may therefore have several coefficients of autocorrelation for a data series Y_t for each k. Because any time k increases by one, we lose another observation on Y_{t+k}, the maximal useful value of k is much less than π , a rough rule is to use $k \leq n/4$, where n is the sample size. An analysis of the patterns of autocorrelation in a data series determines an ARIMA model. To obtain the population autocorrelation coefficient at different lags $k = 1, 2, 3, ..., k$, we use sample data. The explanation of this theoretical coefficient is as

$$\rho_k = \text{cov}(Y_t,\ Y_{t+k}) / \sigma_y^2$$

where, σ_y^2 is the population variance that is formulated as the expected value of $(Y_t - \mu_y)^2$. The population mean of Y_t is the expected value that is denoted as μ_y .

$\mu_y = E(Y_t)$; and $Cov(Y_t, Y_{t+k}) = E[(Y_t - \mu_y)(Y_{t+k} - \mu_y)]$. For a stationary series $Cov(Y_t, Y_{t+k})$, and consequently ρ_k are dependent only on k, the number of time breaks splitting Y_t and Y_{t+k}.

The sample autocorrelation coefficient provides an estimate of ρ_k, is generally calculated as:

$$\hat{\rho}_k = \frac{\sum_{t=1}^{n-k}(Y_t - \overline{Y})(Y_{t+k} - \overline{Y})}{\sum_{t=1}^{n}(Y_t - \overline{Y})^2}$$

The sample autocorrelation function, abbreviated as ACF, is the corresponding set of values. Jenkins and Watts (1968) explored its formulas. Any ρ_k is just a sample value that may vary from zero because of sampling variance. By comparing it with the standard error, we can get an understanding of the magnitude of the sample statistic. An approximate standard error for $\hat{\rho}_k$, computed by Bartlett (1946) is

$$S(\hat{\rho}_k) = \frac{\left(1 + 2\sum_{j=1}^{k-1} \hat{\rho}_j^2\right)}{n^{\frac{1}{2}}}.$$

To check a linear relationship in the population between Y_t and Y_{t+k}, we test the null hypothesis as

H_0: $\rho_k = 0$

Against the alternative hypothesis H_1: $\rho_k \neq 0$.

We then calculate the estimated *t* statistics, $t = (\hat{\rho}_k - \rho_k)/s(\hat{\rho}_k)$. The ratio of the statistic $\hat{\rho}_k$ is estimated with its standard error $s(\hat{\rho}_k)$. Since ρ_k is hypothesized to be zero. If *t* is significant at α% (usually 5% or less), we do not accept the null hypothesis.

3.6.3 Partial Autocorrelation Coefficient

The coefficient of partial autocorrelation is another effective measure of autocorrelation for stationary series. Considering the set of *k* regression equations is a way to estimate the coefficients:

"
$$Y_t = C_1 + \phi_{11} Y_{t-1} + e_{1t}$$

$$Y_t = C_2 + \phi_{21} Y_{t-1} + \phi_{22} Y_{t-2} + e_{2t}$$

.

.

.

$$Y_t = C_{k1} + \phi_{k2} Y_{t-1} + \phi_{22} Y_{t-2} + ... + \phi_{kk} Y_{t-k} + e_{kt}$$
"

In each equation, the population partial autocorrelation coefficients at lag *k = 1, 2, 3, ..., k* are $\phi_{11}, \phi_{22}, \phi_{33}, ..., \phi_{kk}$. Each population coefficient is determined by its sample counterpart ($\hat{\phi}_{kk}$) for the given data set. The subsequent set of values is the sample partial autocorrelation function abbreviated PACF. In calculating $\hat{\rho}_k$, we considered only two random variables Y_t and Y_{t+k}, and we disregarded the intervening random variables Y_{t+k-1}, Y_{t+k-2}, ..., Y_{t-1}. But the role of these random variables in computing $\hat{\phi}_{kk}$ is simultaneously taken into account. We can calculate the importance of each by comparing it with the standard error,

$$S(\hat{\phi}_{kk}) = \frac{1}{\sqrt{n}}.$$

It is suitable to present the PACF for the set of estimates of the $\hat{\phi}_{kk}$ values for $k = 1, 2, 3, ..., K$.

3.6.4 Test of Stationary

The upward or downward trend in the line graph of a time series indicates non-stationary. Remaining in a constant level provides a constant mean, which is an indication of stationary. But this is not a perfect way to test the stationary of a data series. Sometimes a series can be non-stationary in the mean without showing a persistent upward or downward. Several procedures have been proposed to test stationary of a time series.

3.6.4.1 Test of Stationary based on Correlogram

One sample test of stationary is based on the time series so called autocorrelation function (ACF). The ACF at lag k presented by ρ_k is defined as

$$\rho_k = \frac{\gamma_k}{\gamma_0};\ -1 < \rho_k < 1$$

$$= \frac{\text{Covariance at lag } k}{\text{Variace}}.$$

Now if we draw by plotting ρ_k against k, the graph we obtain is known as the population correlogram. Since in training we only have an understanding of a stochastic process. We can only calculate the sample autocorrelation function $\hat{\rho}_k$.

To compute this we must first compute the sample covariance at lag k, $\hat{\gamma}_k$ and the sample variance $\hat{\gamma}_0$ which are defined as

$$\hat{\gamma}_k = \frac{1}{n}\sum_{t=1}^{n}(Y_t - \overline{Y})(Y_{t+k} - \overline{Y}) \qquad \text{and}$$

$$\gamma_0 = \frac{1}{n}\sum_{t=1}^{n}(Y_t - \overline{Y})^2$$

where, n is sample size and $\overline{Y}$ is the sample mean. Therefore the sample auto-correlation function at lag k is $\hat{\rho}_k = \frac{\hat{\gamma}_k}{\hat{\gamma}_0}$. We have defined autocorrelation ρ_k correlograms and their sample counterparts $\hat{\rho}_k$. The statistical significance of any ρ_k can be arbitrated by its standard error. Bartlett has shown that if a time series is purely random that exhibits white noise, the sample autocorrelation coefficient are approximately normally distributed with zero mean and variance $\frac{1}{n}$. Where n is the sample size, so that the standard error of $\hat{\rho}_k$ is $1/\sqrt{n}$. Following the properties of standard normal distribution, the 95% confidence interval of any ρ_k will be ±1.96($1/\sqrt{n}$). If an estimated ρ_k falls inside this confidence interval, we accept the null hypothesis that ρ_k is zero. But if it lies outside the interval we reject the null hypothesis that the true ρ_k is zero. Any significant value of ρ_k breaks the stationary assumption of the data series. If the mean of a series is stationary, then the ACF and PACF will tend to decay quickly toward zero.

3.6.4.2 Complete Significance Test of Autocorrelation

To check the joint hypothesis that all the autocorrelation coefficients ρ_k's are concurrently equal to zero, the Q statistic developed by Box and Pierce, which is defined as

$$Q = n\sum_{k-1}^{m}\hat{\rho}_k$$

where, n is the sample size and m is the lag length. The Q statistics are roughly distributed as a chi-square distribution with m degrees of freedom (i.e. for large samples). If the calculated Q exceeds the critical Q value of the chi-square table at the particular level of significance, we may reject the null hypothesis that all ρ_k are zero in favour of at least some of them must be non-zero. The Ljung Box (LB) statistic is alternatively a Box Pierce Q-statistic, conceived as

$$LB = n(n+2)\sum_{k=1}^{m}\left(\frac{\hat{\rho}_k^2}{n-k}\right) \sim \chi_m^2 .$$

Since both the Q and LB statistics in the broad sample adopt the chi-square distribution with *m* d.f. It is observed that the LB statistic has better small sample properties than the Q-statistic. If the Q-statistic sample value reaches the critical value of χ^2 with *m* degrees of freedom, then at the stated level of significance, at least one value of ρ_k is statistically different from zero. As a key statistic, the Box-Pierce and Ljung Box Q-statistics help to test if the residuals from an estimated ARIMA (p,d,q) model function as a white noise process. Nevertheless, the degrees of freedom are decreased by the number of estimated coefficients when the *m* correlations are formed from the ARIMA (p,d,q) model. Therefore, using the residuals of an ARIMA(p,d,q) model, Q has a degree of freedom of χ^2 with *(m-p-q)* degrees. If a constant term is used, the degrees of freedom are *(m-p-q-1)*.

3.6.5 Unit Root Test

It is possible to describe time series in several respects. Firstly, in the time series, we want to concentrate on the presence of trends. There are two forms of trends: (i) deterministic trends and (ii) stochastic trends. A random walk, which may or may not involve deterministic or stochastic trend, is a stochastic pattern. A time series containing a random walk is referred to as a unit root process.

3.6.5.1 Dickey-Fuller Test

The Dickey-Fuller test checks that a unit root exists in an autoregressive model. It is named after the researchers Dickey and Fuller (1979) who formed the test. Considering the following model is the easiest way to execute this test:

$$Y_t = Y_{t-1} + \varepsilon_t \qquad (3.27.1)$$

where, ε_t is the random error that holds the basic assumptions, it has zero mean and constant variance. The error terms are also not auto-correlated. Such an error term is often referred to as a white noise error. Equation (3.27.1) is a first-order or AR(1) regression in which the value of Y at time t is regressed to its value at time (t-1). If the coefficient of Y_{t-1} is equal to 1, we face the dilemma of the unit root, i.e., a non-stationary condition. Therefore, the regression equation is defined as follows:

$$Y_t = \rho Y_{t-1} + \varepsilon_t \qquad (3.27.2)$$

and find that ρ =1, so we assume that there is a unit root for the stochastic variable Y_t. A time series that has a unit root is regarded as a random walk. The non-stationary time series is a case of random walk.

Alternatively, Equation (4.27.2) is also expressed as

$$\nabla Y_t = (\rho - 1)Y_{t-1} + \varepsilon_t$$

$$\Rightarrow \nabla Y_t = \delta Y_{t-1} + \varepsilon_t \qquad (3.27.3)$$

where, $\delta = (\rho - 1)$ and ∇ is known as the first difference operator. Note that $\nabla Y_t = (Y_t - Y_{t-1}) = \varepsilon_t$.

Making use of the definition, we can easily see that (3.27.2) and (3.27.3) are the same. Nevertheless, now the null hypothesis is that δ =0

if δ is in fact 0, we can write (3.27.3) as

$$\nabla \mathrm{Y}_t = (\mathrm{Y}_t - \mathrm{Y}_{t-1}) = \varepsilon_t \qquad (3.27.4)$$

where, (3.27.4) states that first difference in a random walk time series (= ε_t) is stationary. To find out a time series Y_t is non-stationary, run the regression (3.27.2) and test $\hat{\rho} = 1$, or estimate (3.27.3) and check $\hat{\delta} = 0$ based on the t-statistic. Unfortunately, even in large samples, it does not follow t-distribution. Under the null hypothesis that ρ = 1, the conventional t statistics are referred to as τ (tau) statistics whose critical values were tabulated according to Monte Carlo simulations that were developed by Dickey and Fuller.

The tau test is recognized by researchers as the Dickey-Fuller (DF) test. The conventional t test is used to test the null hypothesis $\rho = 1$. If the null hypothesis $\rho = 1$ is rejected, then the time series is regarded as stationary. If the estimated absolute value of the τ statistic surpasses the critical values of the DF or Dickey-Fuller absolute values, then we accept the null hypothesis i.e., the provided time series is stationary. The time series is non-stationary if it is smaller than the critical value. The Dickey-Fuller test applies to regressions that use the following forms for theoretical and functional reasons:

$$\nabla Y_t = \delta Y_{t-1} + \varepsilon_t \quad (3.27.5)$$

$$\nabla Y_t = \beta_1 + \delta Y_{t-1} + \varepsilon_t \quad (3.27.6)$$

$$\nabla Y_t = \beta_1 + \beta_2 t + \delta Y_{t-1} + \varepsilon_t \quad (3.27.7)$$

where, t is the time or trend variable. In every case the null hypothesis is that $\delta = 0$. Then, there is a unit root. The difference between (3.27.5) and the other two regressions lies in the presence of the constant (intercept) and the trend term.

3.6.5.2 Augmented Dickey-Fuller Test

Dickey and Fuller (1979) developed the Augmented Dickey-Fuller (ADF) test. It is the most commonly used test in a time series to check the unit root property. Three distinct regression equations that can be used to check the existence of unit root in a time series are as follows:

$$\nabla Y_t = \delta Y_{t-1} + \sum_{j=1}^{J} \varphi_j \nabla Y_{t-j} + \mu + \beta t + \varepsilon_t \quad (3.27.8),$$

$$\nabla Y_t = \delta Y_{t-1} + \sum_{j=1}^{J} \varphi_j \nabla Y_{t-j} + \mu + \varepsilon_t \quad (3.27.9) \text{ and}$$

$$\nabla Y_t = \delta Y_{t-1} + \sum_{j=1}^{J} \varphi_j \nabla Y_{t-j} + \varepsilon_t \quad (3.27.10).$$

The dependent variable in the equation (3.27.8) and (3.27.10) is ∇Y_t. This means that if $\delta = 0$, Y_t has a unit root. Hence, the null hypothesis of the test equations

(3.27.8) and (3.27.10) states the non-stationary of Y_t : H_0: $\delta = 0$ (Y_t has a unit root). The deterministic regressors vary from the three equations. A major issue in unit root testing is the alternative for the three equations. One concern is that the estimated additional parameters minimize the degree of freedom and the power of the test. Reduced power means that the researcher will assume that, where it is not the case, the process has a unit root. The second issue is that a suitable test statistic $\delta = 0$ depends on which regressors are used in the equation. For instance, if a deterministic term is used in the data-generating process, omitting the term β_t' gives an upward bias in the expected value. Additional regressors, nevertheless, raise the absolute value of the critical values such that the null hypothesis of a unit root cannot be rejected by the researcher. The test is carried out by way of the usual t- statistics of $\hat{\delta}$. The t-statistics of the three models are denoted as t_τ, t_μ and t, respectively. F-statistics were proposed by Dickey and Fuller (1979) to measure the joint hypotheses $\delta=\beta=\mu=0$ (Φ) and $\delta = \beta = 0$ (Φ_3) in equation (3.27.8) and the joint hypothesis $\delta = \mu = 0$ in equation (3.27.9), denoted as Φ_1. The t-statistics and t-tau and t-mu and the F-statistics Φ_2 and Φ_3 do not have the standard t and F-distributions under the assumption of non-stationary but are functions of Brownian motions. Critical values of the asymptotic distributions of such t-statistics were given by Fuller (1979). MacKinnon (1996) enhanced them across the larger sets of repetitions. For the F- statistics of Φ_1, Φ_2, and Φ_3, Dickey and Fuller (1979) described the critical values. As illustrated in Figure 3.5, Dolado et al. (1990) established a rigorous testing approach between the alternative equations. There are several phases in the method of unit root testing:

Step 1. The null hypothesis of stationary is tested with t-tau in the most unregulated equation (3.27.8). The time series Y_t is stationary in trend if the null hypothesis is rejected and there is no need for further progression.

Step 2. If the null hypothesis is not rejected, we use the statistics Φ_3 to measure the validity of the deterministic trend under the null hypothesis $\delta = \beta = 0$. If it is significant, it is necessary to assess the further existence of the unit root, noticing that the t-statistic now follows a standard t-distribution.

Step 3. In the equation (3.27.8), δ and β are jointly insignificant, we approximate the equation without the deterministic trend (equation (3.27.9)) and the unit root test using t-mu and its critical values. We will stop again if the null hypothesis is rejected, and assume that the variable Y_t is stationary.

Step 4. If the null hypothesis is accepted, we test under the null $\delta = \mu = 0$ using Φ_1 for the validity of the constant term. We evaluate the unit root test using the standard normal distribution if the constant term is significant.

Step 5. In Equation (3.27.9), if δ and μ are mutually insignificant, we calculate equation (3.27.10) and test for the existence of a unit root. The method either ends with the consequence that the variable Y_t is stationary or that a unit root is included in Y_t. If in each of the steps of the strategy, we do not dismiss the null hypothesis, we infer that Y_t is non-stationary and needs to be differenced at least one to become stationary. We start by checking the differenced series to detect the order of integration *d* of the time series Y_t until the unit root assumption is dismissed. Therefore, if Y_t is discovered to be non-stationary and ∇Y_t is revealed to be stationary, Y_t is called 'integrated of order 1' (referred to as $Y_t \sim I(1)$). If, after differentiating *d* times, we can only deny the null of a unit root, we infer that the series is integrated of order *d*. Stochastic patterns are sometimes linear and often quadratic, so *d* is hardly ever greater than 2 (Leeflang et al. 2000). The number of lags in equations (3.27.8) and (3.27.10) is evaluated by selection criteria AIC, BIC that was proposed by Phillips and Perron (1988). The ADF test implies that the variable being considered is continuous and that certain real values can be occupied.

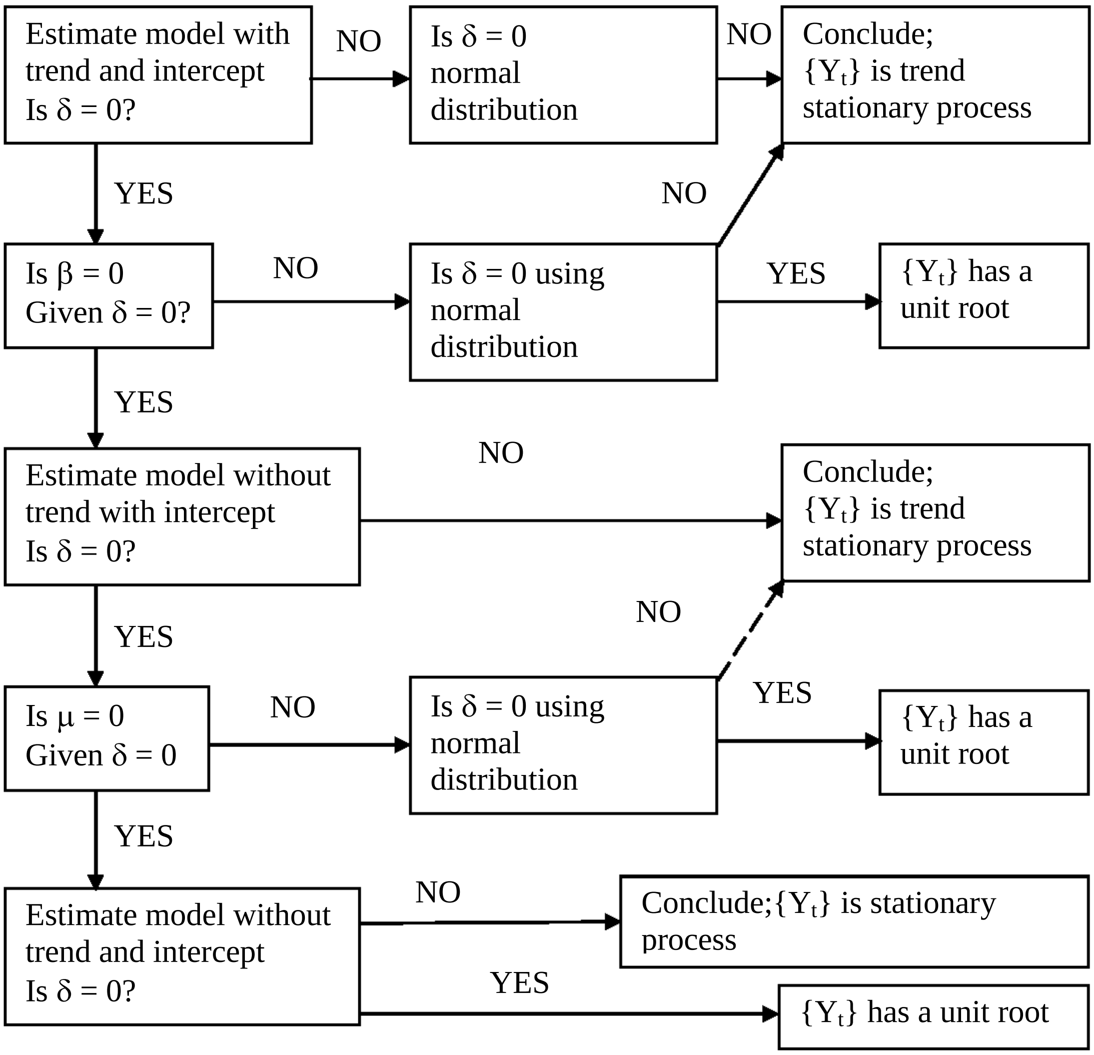


Figure 3.5: Functional diagrams of a systematic strategy of unit root test

3.6.6 Normality Checking

A popular assumption is that random residuals are normally distributed. This allows us to run t-tests on coefficient significance at the estimation stage. The study of the residual histogram is one method of testing normality. Another is a normal residual probability plot (Cook and Weisberg, 2009; Liu and Hudak, 1986). A helpful graphical presentation of the data is offered by both procedures; however, they did not include any formal test of normality. Different formal tests exist for normality. In the following subsection, the most recent and powerful test for the normality of the residuals is discussed.

3.6.6.1 Jarque Bera Test for Normality

We assumed the normality of the error term of the regression model. In finite samples, the hypothesis testing and confidence interval estimation of the parameters of the regression equation relied on the normality assumption. The assumption of normality of errors (residuals) can be tested by the Jarque-Bera (JB) test (1987). JB normality test is two degrees of freedom χ^2 test based on the skewness and kurtosis coefficients. For a normal distribution skewness = 0 and kurtosis = 3. Many computer programs compute these two statistics. Some computer programs compute excess kurtosis instead of kurtosis. From sample data, one can compute these as follows:

Suppose the data is x. Compute sample mean $\bar{x}$ and compute the following:

$$m_2 = \left(\frac{1}{T}\right)\sum (x_t - \bar{x})^2, \qquad m_3 = \left(\frac{1}{T}\right)\sum (x_t - \bar{x})^3, \qquad m_4 = \left(\frac{1}{T}\right)\sum (x_t - \bar{x})^4$$

Skewness (Sk.) $= \dfrac{m_3}{(m_2)^{3/2}}$ and Kurtosis (ku.) $= \dfrac{m_4}{(m_2)^2}$.

The Jarque Bera (JB) statistic takes the following form:

$$\mathrm{JB} = \mathrm{T}\left(\frac{(\mathrm{sk})^2}{6} + \frac{(\mathrm{ku}-3)^2}{24}\right), \ \mathrm{JB} \sim \chi^2(2).$$

The observed value of JB is compared to the critical value of χ^2 (2) and the test is concluded.

3.6.7 Checking Outliers

Outliers are abnormal observations that are distinct from the majority. These might even be a part of the data or might be due to gross errors such as inappropriate key punching. The latter type is, of course, conveniently fixed. Nevertheless, significant outliers are more difficult to control. Deneshkumar and Sentham (2011) anticipated that the undetected outliers can impact any consequent review of the data set. A careful inspection of the time series plot is necessary to identify whether there are outliers. In order to detect the presence of outliers, standardized residual plots are also suitable. The data for which we get the standardized residuals outside of the range (-3, 3) are considered as outliers.

3.7 Forecasting Algorithm

Assuming that adequate historical data is usable, the forecasting algorithms will then be processed to train the models with approximately 75% samples and predict the next 25% observations for testing. Using a suitable criterion, compare the forecasts to the actual values. To perform the required out-of-sample forecasts, the forecasting techniques (models) that provide the smallest value of RMSE for the test set on the original data set are applied.

3.8 Software Used

The advancement of computer and information technology makes the procedure of analyzing data easier. To explore the data quickly, easily, and accurately, there is no alternative to computer programs and software. Various computer programs and software have been used to complete this study. During the preparation of this dissertation, Microsoft Word, Microsoft Excel, Eviews, STATISTICA, and R software were used.

Chapter Four
Results and Discussion

This chapter discusses the results of this dissertation. It has applied the methodology that includes the proposed support system and proper models of the selected indicators of DSE. The results are described based on the findings of relevant data. The crucial models and forecasting metrics are computed that make a comparison to approaches of other studies.

4. Exploratory Data Analysis

Exploratory Data Analysis (EDA) discloses the characteristics of data and patterns of data analysis. It is a vigorous method of graphical representation of data series. It exposes the results of the influence of unusual data values.

4.1 Time Series Plot

The most common and widely used EDA technique is the time series plot. Researchers and econometricians are interested to see the graphical pattern of collecting data. In a time series plot, data are plotted against their occurrence of time. The vertical axis and horizontal axis denote the value and observation time of variables, respectively. If the time series plot shows a strong up or down trend, we may conclude that the data series is non-stationary. In the case of non-stationary data, the seasonal or non-seasonal differencing transformations are used to make the series stationary. Data series may be non-stationary sometimes without displaying any upward or downward trend.

4.1.1 Time Series Plot of Microeconomic Indicators

Based on selected microeconomic indicators of Bangladesh, this study analyzes the effect of DSE portfolios. The microeconomic indicators are the number of TEC and IMC in US$ which have a direct and immediate impact on STR of DSE.

The time series plot of microeconomic indicators of DSE like STR, TEC, and IMC from 1990 to 2012 is shown in Figure 4.1.

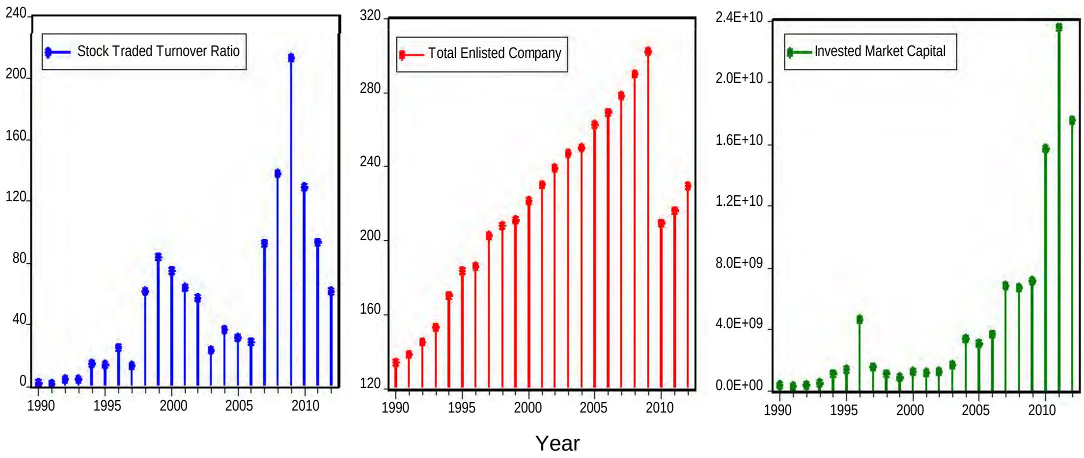


Figure 4.1: Time series plot of STR, TEC, and IMC

Figure 4.1 illustrates that there are upward trends of STR, TEC, and IMC. The rising scenario in SRT was in 1999 and 2009; the number of TEC was gradually increasing from 1990 to 2009 and then gradually up and down after 2009.

4.1.2 Time Series Plot of Macro Economic Indicators

The macroeconomic indicators are GDP, GNI, GS, GI, DIR, and GFI which have the indirect and long-run impact on DSE portfolios. The time series plot of macroeconomic indicators like GDP, GNI, GS, GI, DIR, and GFI of DSE along with DGI from 2005 to 2012 is shown in Figure 4.2. In Figure 4.2, it is reported that DGI peaked in 2011 and that the downward trend occurred after 2005, 2008, and 2011; GDP, GNI, and GS were gradually increasing, and at the same time there were varying figures in GI, DIR, and GFI.

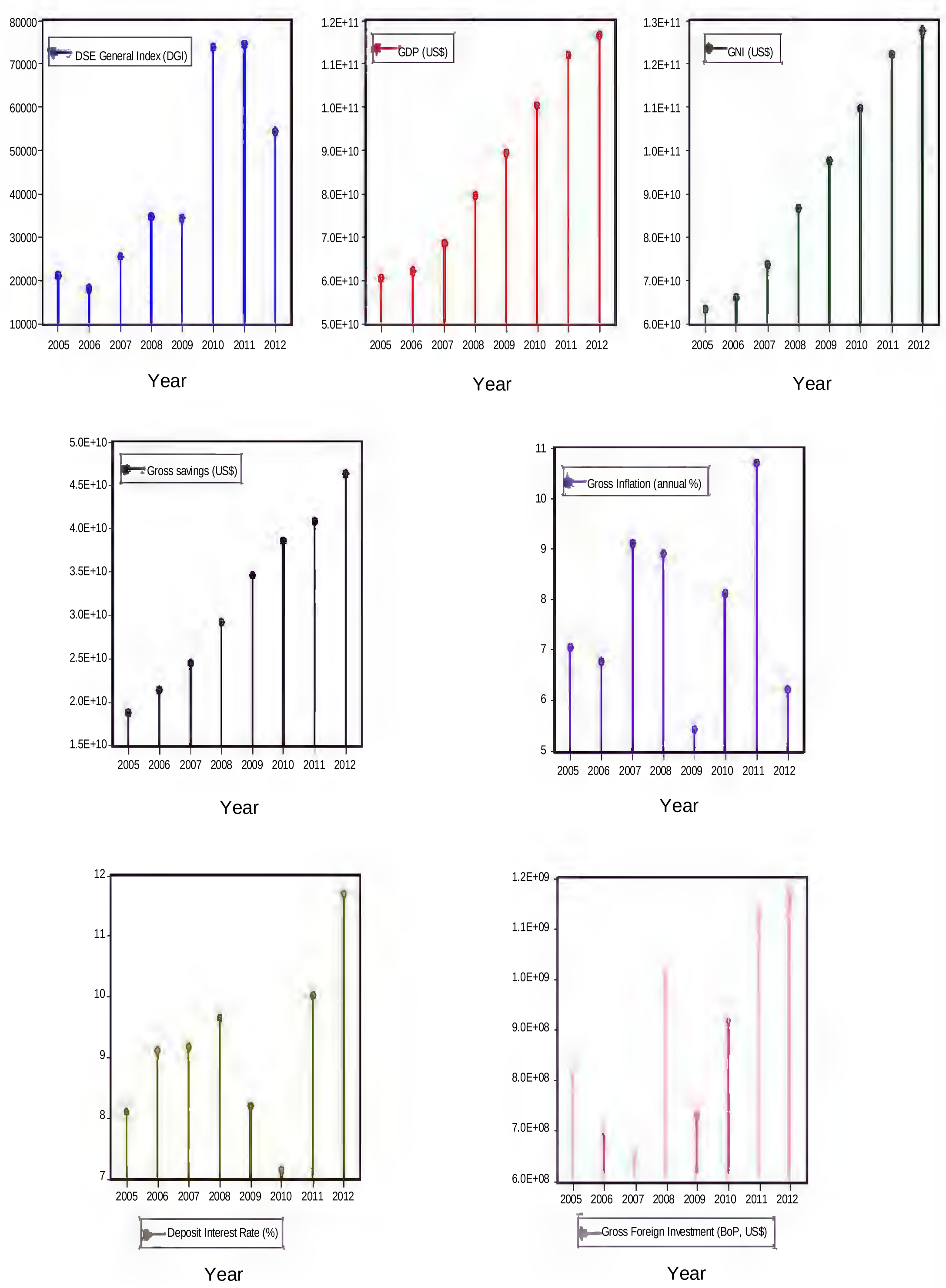
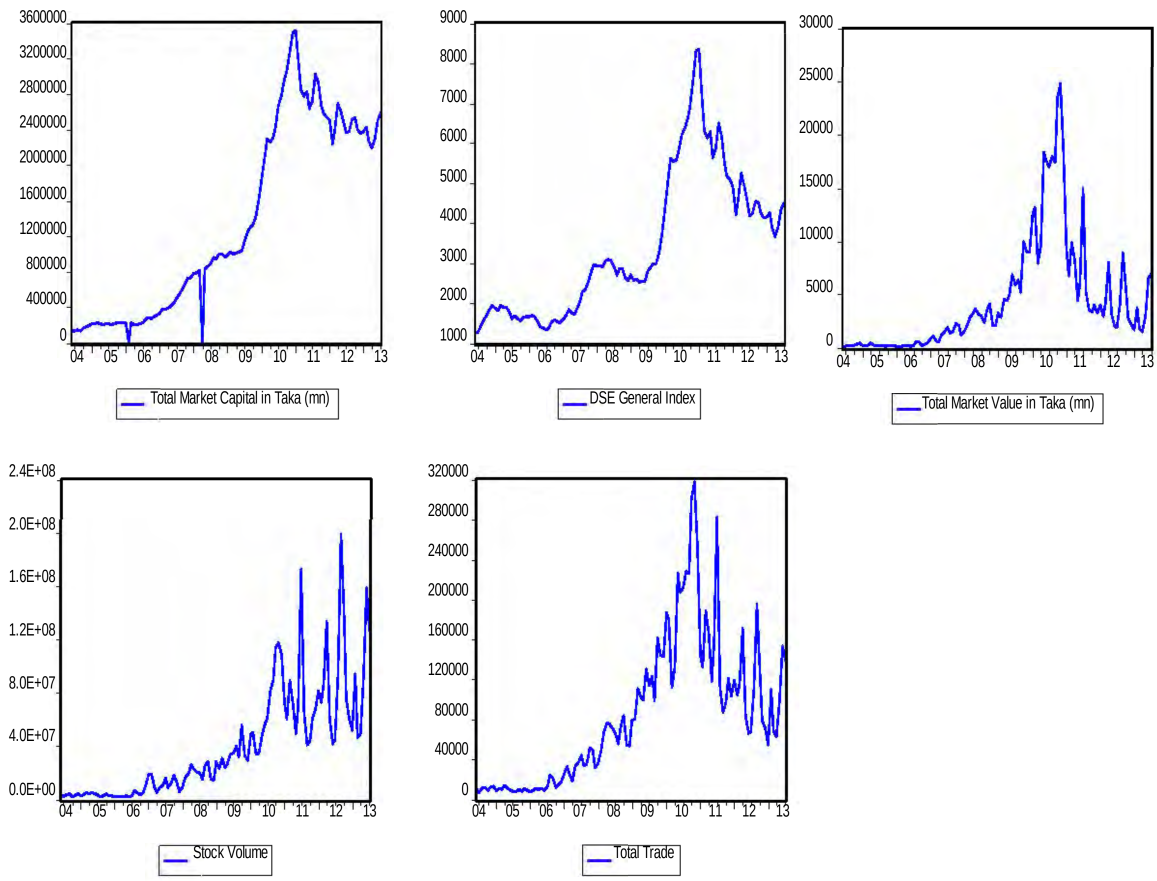


Figure 4.2: Time series plot of DGI, GDP, GNI, GS, GI, DIR, and GFI

The time series plots of total invested stock market capital in Taka (mn), DGI, stock trade, stock volume, and current market value in Taka (mn) for the period of June 2004 to July 2013 are shown in Figure 4.3. In Figure 4.3, it is reported that each series rose from July 2010 except for stock volume. In addition, there were significant volatilities in each series from 2010 until the end of the trading day.



Note: Years are plotted on the horizontal axis and the respective data points are plotted on the vertical axis.

Figure 4.3: The time series plot of stock market capital, general index, stock trade, stock volume, and current market value of DSE

4.1.3 Time Series Plot of Newest Indicators

The newest indicators of DSE are as DSEX index, DSES index, and DSE30 index from January 2014 to December 2018 on a monthly scale. The time series plot of the newest indicators of DSE from January 2014 to December 2018 is shown in Figure 4.4.

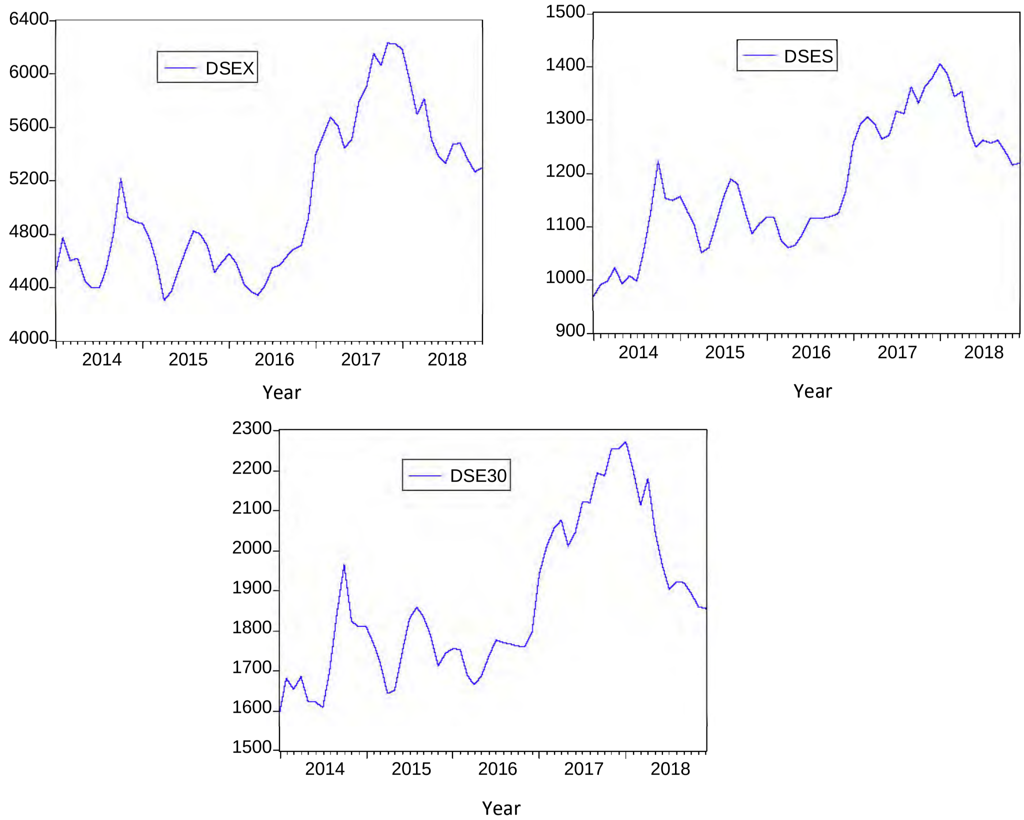


Figure 4.4: Time series plot of the newest indicators of DSE

As shown in Figure 4.4, there is an upward trend for DSEX, DSES, and DSE30 indices, respectively, and there is no seasonality in DSEX, DSES, and DSE30 indices. In contrast, after the month of July 2017, DSEX, DSES, and DSE30 indices gradually declined. However, regular cyclical fluctuations were not found in DSEX, DSES, and DSE30 indices.

4.2 Cobb-Douglas (CD) Functional Regression Model

The Cobb-Douglas (CD) functional form of production functions is widely used in economics to describe the relationship between output and input. Despite many other factors affecting economic progress, the CD model has proven to be remarkably accurate. This nonlinear function is used to construct a suitable model for predicting the stock turnover ratio.

4.2.1 Estimation of CD Functional Regression Model

To investigate the direct and immediate impact on the portfolios of DSE prices, the CD functional regression form is used considering the output level STR as a dependent variable and the IMC and TEC of DSE as the independent variables. To check the existence of multicollinearity between the explanatory variables—IMC and TEC, correlation analysis is conducted. The result of the correlation analysis is reported in Table 4.1. Multicollinearity problem has not been found because none of the correlation coefficients between IMC and TEC is greater than 0.80.

Table 4.1: Pearson coefficient of correlation matrix between IMC and TEC

Variable	IMC	TEC
IMC	1	0.275
TEC	0.275	1

The estimation results of CD functional regression of STR on the explanatory variables IMC and TEC using by OLS method are reported in Table 4.2.

Table 4.2: The estimation results of CD functional regression model

Variables	Coefficient	Std. Error	t-Statistic	Prob.
C	-25.00805	3.490276	-7.165063	0.0000
Log(IMC)	0.363238	0.160425	2.264228	0.0348
Log(TEC)	3.842619	0.881953	4.356942	0.0003
R-squared value	0.775131	F-statistic		34.47036
Log likelihood	-23.34067	Prob.(F-statistic)		0.000000

Dependent variable: Log (STR) Sample range: 1990 to 2012 (annual)
Data source: http://data.worldbank.org/country/bangladesh

The estimation of CD functional regression of STR (Y) on IMC (k) and TEC (L) by OLS gives the following equation: LogY = -25.00805+ 3.842619LogL+

0.363238LogK. The intercept and slope coefficients of all explanatory variables are statistically significant at least at the 5% level. There found that total factor turnover ratio, c = -25.00805 which is negative, therefore there was an overall 25 point negative STR due to the fixed cause and the relationship of STR with TEC is positive (3.842619) and with IMC is also positive (0.363238) over the period 1990 to 2012. This result implies that in DSE if 100 points increase TEC then STR also may increase 384.2619 points and if 100 points increase IMC then STR also may increase 36.3238 points. Moreover, F-statistic = 90.02 and Prob. value = 0.000 imply that the regression model significantly fits the data. Finally, the R-square value indicates that about 77.5131 percent variations of STR are explained by the explanatory variables—IMC and TEC of DSE. The estimation results of this CD functional regression of the actual, fitted, and residual plot is shown in Figure 4.5 and the actual, fitted, and residual value with residual plot is also shown in Figure 4.6. From Figure 4.5 and Figure 4.6, in the financial year 1991, 2005, and 2006, the fitted values of STR are over the actual value and exceed outside the confidence interval at 5% level, and in 1998, 1999, 2000, and 2010 financial year the fitted values of STR are below the actual value and exceed outside the confidence interval at 5% level and in 1990, 1992, 1993, 1994, 1995, 1996, 1997, 2001, 2002, 2003, 2004, 2005, 2006, 2007, 2009, 2011, 2012 financial year, the fitted values of STR lie in the confidence interval at 5% level.

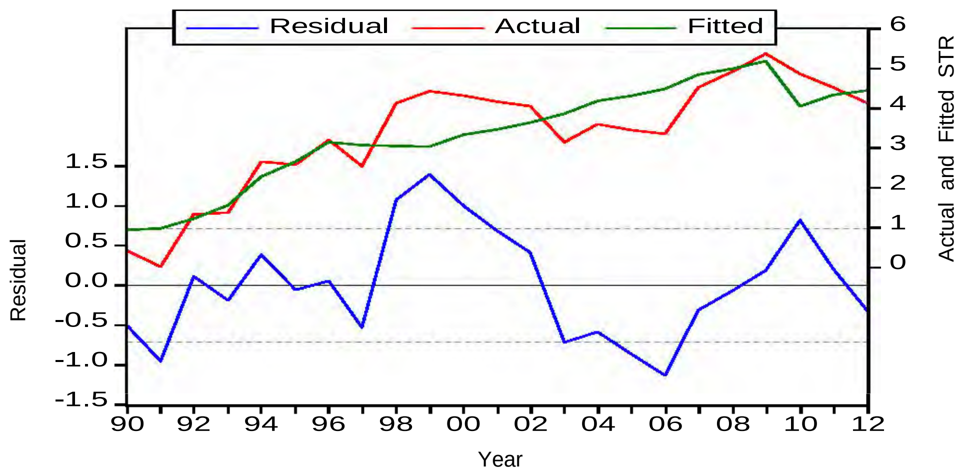


Figure 4.5: Actual, fitted, and residual plot of STR from CD model

obs	Actual	Fitted	Residual
1990	0.40922	0.92721	-0.51799
1991	0.01681	0.97604	-0.95923
1992	1.32803	1.22236	0.10566
1993	1.36388	1.56185	-0.19797
1994	2.65593	2.27207	0.38386
1995	2.58270	2.64327	-0.06057
1996	3.19950	3.15042	0.04908
1997	2.53749	3.07321	-0.53573
1998	4.11703	3.04170	1.07533
1999	4.41915	3.03207	1.38708
2000	4.31528	3.32446	0.99082
2001	4.15262	3.46494	0.68768
2002	4.04216	3.62747	0.41468
2003	3.14616	3.86549	-0.71933
2004	3.58504	4.17184	-0.58681
2005	3.44992	4.31978	-0.86986
2006	3.34558	4.48409	-1.13851
2007	4.52493	4.84017	-0.31524
2008	4.92187	4.99594	-0.07407
2009	5.35923	5.17275	0.18648
2010	4.86108	4.04783	0.81325
2011	4.52802	4.32203	0.20599
2012	4.11376	4.43838	-0.32462

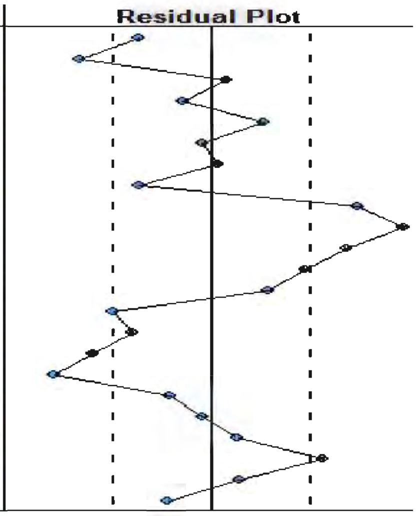


Figure 4.6: Actual, fitted, and residual value of CD model of STR on IMC and TEC

4.2.2 Multicollinearity Diagnostics

Multicollinearity refers to the presence of a perfect or exact linear relationship among the explanatory variables of a regression model. Prediction and forecasting are the main purposes of regression analysis. In regression analysis, multicollinearity is not a severe problem to forecast a time series better due to the higher value of R^2. Besides, the objective of the study is not only forecasting but also estimating the parameters reliably. So, severe multicollinearity will be a problem for large standard errors of the estimators. The multicollinearity of CD functional regression analysis is checked by Variance Inflation Factor (VIF) and Tolerance Value (TV).

4.2.2.1 Variance Inflation Factor (VIF)

One of the diagnostics of multicollinearity is the Variance Inflation Factor (VIF). Its threshold is 10. When VIF is less than 5, no multicollinearity; $5 \leq \text{VIF} \leq 10$, moderate multicollinearity and larger than 10, the impact of multicollinearity is strong. Table 4.3 shows VIF statistics between IMC and TEC. VIF statistics between IMC and TEC is 1.841 which is less than 5. Therefore, there is no presence of multicollinearity between IMC and TEC.

4.2.2.2 Tolerance Value (TV)

Tolerance Values (TV) are defined as the inverse of VIF. When TV is greater than 0.2, no multicollinearity; $0.1 \leq TV \leq 0.2$, moderate multicollinearity, and less than 0.1, the impact of multicollinearity is severe. Table 4.3 shows TV statistics between IMC and TEC. TV statistics between IMC and TEC is 0.543 which is greater than 0.2. Therefore, there is no presence of multicollinearity between IMC and TEC.

Table 4.3: Multicollinearity diagnostics using VIF and TV

Explanatory Variables	Collinearity Statistics	
	VIF	**TV**
IMC	1.841	0.543
TEC	1.841	0.543

Dependent variable: STR

4.2.3 Residuals Normality Checking

The value of the Jarque-Bera test statistic for the residual of the CD functional regression model is 0.814 with significant probability (P-value = 0.665) and the Anderson Darling test statistic is 0.166 with significant probability (P-value = 0.930). These tests suggest that the null hypothesis of residuals- CD functional regression model does not come from a normal distribution is rejected at the 5% level of significance. The normal probability plot is reported in Figure 4.7. The normal probability plot of the estimated residuals is a little S-patterned curve rather than a straight line and it suggests that the estimated residual may be normal.

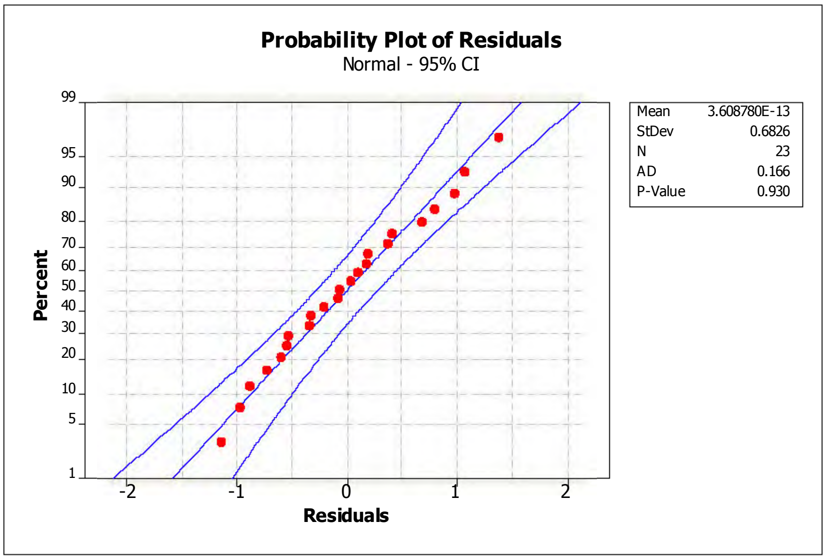


Figure 4.7: Normal probability plot of residuals of CD functional regression model

4.2.4 Outliers Checking

The Standardized residual plot of the CD functional regression model has some positive and negative values that fall in the standard deviations confidence interval except the year 1991, 1998, 1999, 2000, 2005, and 2006 shown in Figure 4.8. The influence of outliers was observed in the years 1991, 1998, 1999, 2000, 2005, and 2006 respectively.

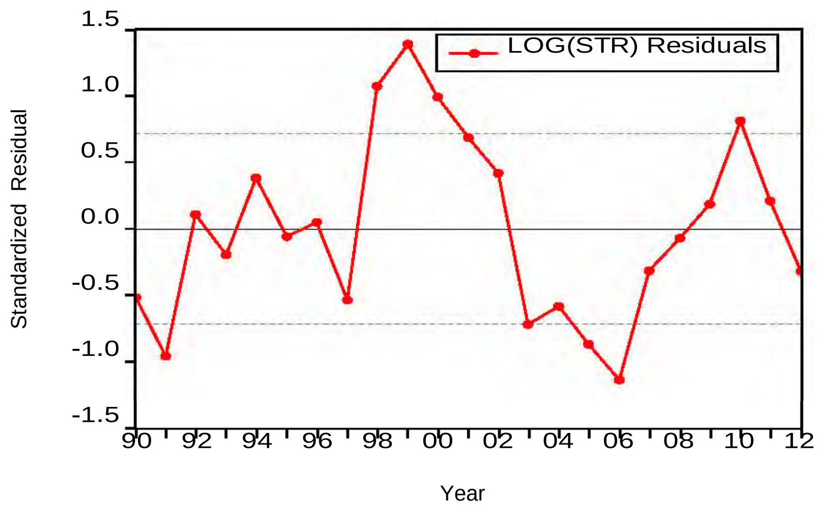


Figure 4.8: Standardized residual plot of CD functional regression model

4.2.5 Wald Hypothesis Testing

The Wald hypothesis of constant returns to scale is then tested as the restriction under H_0: $\alpha+\beta = 1$ which is reported in Table 4.4. From Table 4.4, the null hypothesis is accepted at the 1% level of significance. Therefore, the elasticity of STR in DSE with respect to IMC and TEC is a constant return to scale.

Table 4.4: Wald hypothesis test summary of constant returns to scale

Test Statistic	Value	df	Probability
Chi-square	16.785	1	0.000
Null hypothesis summary:			
Normalized restriction (=0)		Value	Std. Err.
$-1 + \alpha+\beta$		3.206	0.783

Restrictions are linear in coefficients.

4.3 Multiple Linear Regression Analysis

To examine the indirect and long-run impact on the portfolios of DSE prices, the multiple log-linear regression model is applied considering the output level DGI as the dependent variable and the macroeconomic indicators like GDP, GNI, GS, GI, DIR, and GFI, respectively as the independent variables. The estimation results of the multiple log-linear models of DGI on the independent variables GDP, GNI, GS, GI, DIR, and GFI are conducted by the OLS method which is reported in Table 4.5. From Table 4.5, it is found that the total factor of DGI, $\alpha = -44.93569$ which is negative and significant at 5% level. The coefficients of Log(GDP), Log(GNI), and Log(GFI) are significant at 10% level and the coefficients of Log(GI) and Log(DIR) are significant at 5% level and Log(GS) is still insignificant at 10% level. GDP, GS, GI, and GFI have a positive impact, and GNI and DIR have a negative impact on DGI. Finally, the R-squared value indicates that about 99.9889 percent variation of DGI is explained by total variations among independent variables GDP, GNI, GS, GI, DIR, and GFI, respectively. Moreover, F-statistic = 1502.362 and Prob. value = 0.019746 imply that the regression model significantly fits the data at the 5% level.

Table 4.5: The estimation results of multiple log-linear regression model of DGI on GDP, GNI, GS, GI, DIR, and GFI, respectively

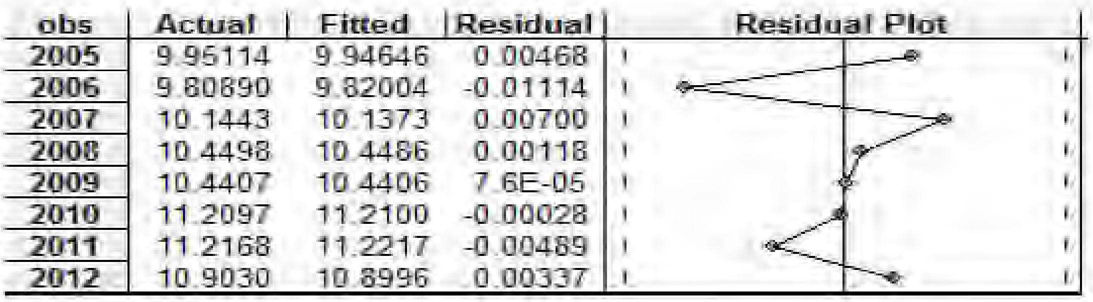
Variable	Coefficient	Std. Error	t-Statistic	Prob.
α (Constant)	-44.93569	1.240996	-36.20938	0.0176
Log(GDP)	18.18548	2.557990	7.109285	0.0890
Log(GNI)	-19.46508	3.050490	-6.380970	0.0990
Log(GS)	3.264091	0.542459	6.017209	0.1048
Log(GI)	0.753333	0.041428	18.18422	0.0350
Log(DIR)	-1.171250	0.050185	-23.33866	0.0273
Log(GFI)	0.554852	0.066922	8.290988	0.0764
R-squared value	0.999889			
F-statistic	1502.362			
Prob.(F-statistic)	0.019746			

Dependent Variable: Log(DGI)

Sample Range: 2005 to 2012 (Annual)

Data source: GDP, GNI, GS, GI, DIR and GFI from http://data.worldbank.org/country/bangladesh and DGI from http://www.dsebd.org/recent_market_information.php

The estimation of multiple log-linear regression of DGI on GDP, GNI, GS, GI, DIR, and GFI by OLS gives the R-square value 0.999889 and overall model fitting is significant at 5% level (F-statistic= 1502.362, p. value < 0.05). The actual, fitted, and residual value with a residual plot is shown by Figure 4.9 which implies that the fitted values are approximately identical with the actual value and the estimated residuals lie in the 5% confidence interval.



obs	Actual	Fitted	Residual
2005	9.95114	9.94646	0.00468
2006	9.80890	9.82004	-0.01114
2007	10.1443	10.1373	0.00700
2008	10.4498	10.4486	0.00118
2009	10.4407	10.4406	7.6E-05
2010	11.2097	11.2100	-0.00028
2011	11.2168	11.2217	-0.00489
2012	10.9030	10.8996	0.00337

Figure 4.9: The actual, fitted, and residual value with a residual plot of DGI on GDP, GNI, GS, GI, DIR, and GFI

4.3.1 Multicollinearity Diagnostics

The multicollinearity of multiple log-linear regression of DGI on GDP, GNI, GS, GI, DIR, and GFI is checked by Variance Inflation Factor (VIF) and Tolerance Value (TV) reported in Table 4.6. When VIF is 10 or larger, or TV is less than 0.1, the impact of multicollinearity is severe; therefore the variable should be removed or re-estimated the models. GDP, GNI, and GS have a severe impact of multicollinearity.

Table 4.6: Multicollinearity diagnostics using VIF and TV

Explanatory Variables	**Collinearity Statistics**		**Influence of Multicollinearity**
	VIF	TV	
GDP	112.434	0.009	Severe
GNI	20990.675	4.764E-5	Very Severe
GS	92.863	0.011	Severe
GI	1.193	0.839	No
DIR	1.501	0.666	No
GFI	5.100	0.196	Moderate

Dependent variable: DGI

Table 4.6 shows there is very severe multicollinearity for GNI, severe multicollinearity for GDP and GS, and moderate multicollinearity for GFI. So, the multiple linear regression model is re-estimated by dropping GNI due to very severe multicollinearity and standardized GDP, standardized GS and standardized GFI are used as the explanatory variables due to severe/moderate multicollinearity. Model summary of re-estimated multiple linear regression is presented in Table 4.7. From Table 4.7, we conclude that the overall model fitting is significant at a 10% level (F = 14.639, P. value < 0.10) and higher value of R-square (0.973) is found. There are also negative influence of DGI due to GS (standardized coefficient = -0.146) and DIR (standardized coefficient = -0.402), and positive influence of DGI due to GDP (standardized coefficient = 1.038) and GI (standardized coefficient = 0.263), respectively.

Table 4.7: Summary of re-estimated multiple linear regression model of DGI

Particular	**Sum of Squares**	**df**	**Mean Square**	**F**	**Prob.**	**R Square**
Regression	3516741715.645	5	703348343.129	14.639	0.065	0.973
Residual	96091594.084	2	48045797.042			
Total	3612833309.729	7				

Model	**Unstandardized Coefficients**		**Standardized Coefficients**
	B	**Std. Error**	**Beta**
(Constant)	75379.522	27024.769	
GDP	23580.736	31177.318	1.038
GS	-3324.262	28982.483	-0.146
GI	3429.805	1947.220	0.263
DIR	-6578.439	2525.662	-0.402
GFI	3530.883	6080.245	0.155

Dependent Variable: DGI; Predictors: (Constant), GFI, GI, DIR, GS, GDP
Sample Range: 2005 to 2012 (Annual)

4.3.2 Residuals Normality Checking

The value of the Jarque-Bera test statistic for the residual of multiple log-linear regression of DGI on GDP, GNI, GS, GI, DIR, and GFI is 0.890 with significant probability (P. value = 0.640) and Anderson Darling test statistic is 0.301 with significant probability (P. value = 0.498). These tests suggest that the null hypothesis of residuals from multiple Log-linear regression of DGI on GDP, GNI, GS, GI, DIR, and GFI do not come from a normal distribution is rejected at the 5% level of significance. The normal probability plot is reported in Figure 4.10. The normal probability plot of the estimated residuals is a little S-patterned curve rather than a straight line and so it suggests that the estimated residual may be normal.

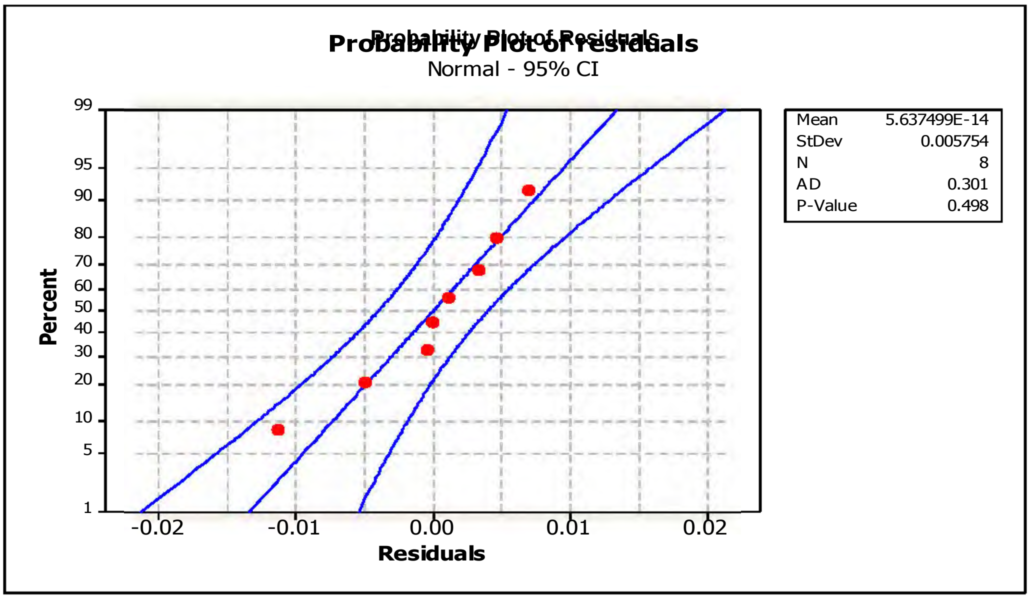


Figure 4.10: Normal probability plot of residuals of multiple log-linear regression model

4.3.3 Outliers Checking

The Standardized residual plot of multiple log-linear regression model of DGI on GDP, GNI, GS, GI, DIR, and GFI has some positive and negative values that fall in the standard deviations confidence interval shown in Figure 4.11. Therefore, the multiple log linear regression model of DGI on GDP, GNI, GS, GI, DIR, and GFI is free from outliers.

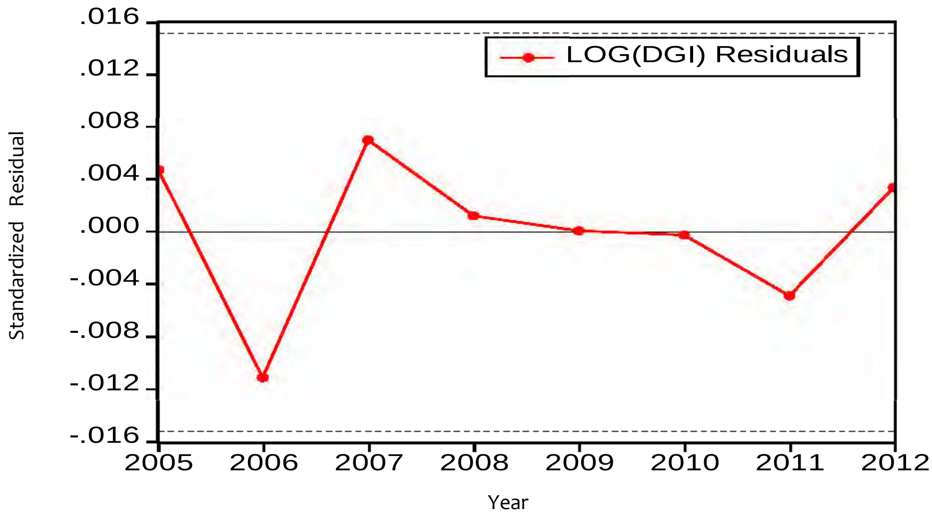


Figure 4.11: Standardized residual plot of multiple log-linear regression model

4.4 VAR Modeling and Forecasting

Selected indicators of DSE in Bangladesh and the microeconomic variables such as invested market capital, DSE General Index (DGI), current market value, stock volume, and stock trade from June 2004 to July 2013 are used as the basis on the daily scale. But to get the maximum explorative information and reduction of volatility, the data have been transformed to the monthly scale. Data from June 2004 to June 2012 are used in training and from July 2012 to July 2013 are used in testing samples for modeling and analyzing purposes. The summary statistics of market capital in Taka (mn), DSE General Index, market value, stock volume, and trade of DSE are shown in Table 4.8.

Table 4.8: Summary statistics of market capital, DGI, value, volume, and trade of DSE

Variable	Statistics	Results
Market Capital in Taka (*mn*)	Mean	1,349,236
	5% trimmed mean	1,311,896
	Median	9,98,774.6
	Std. deviation	1.08E + 06
	Minimum	1,600
	Maximum	3,512,212
	Range	3,510,612
DSE General Index	Mean	3,415.11
	5% trimmed mean	3,298.25
	Median	2,907.92
	Std. deviation	1,812.722
	Minimum	1,270
	Maximum	8,340
	Range	7,070
Value in Taka (*mn*)	Mean	4,395.16
	5% trimmed mean	3,745.74
	Median	2,800.02
	Std. deviation	5,327.383
	Minimum	120
	Maximum	24,827
	Range	24,708
Volume	Mean	3,89,66,424
	5% trimmed mean	3,41,33,017
	Median	2,57,06,199
	Std. deviation	4,12,35,891
	Minimum	16,25,758

Variable	Statistics	Results
Volume (Continued)	Maximum	19,84,40,256
	Range	19,68,14,498
Trade	Mean	83,572.28
	5% trimmed mean	77,348.02
	Median	69,859.33
	Std. deviation	72,565.02
	Minimum	6,427
	Maximum	3,16,926
	Range	3.10,500

Box and Whisker plot is used to investigate the data series of DSE, how much percent of data is representing maximum frequencies; non-outlier range, and how much is affected by outliers and extreme values. The Box and Whisker plot of market capital, general index, value, volume, and trade of DSE, respectively is shown in Figures 4.12, 4.13, 4.14, 4.15, and 4.16.

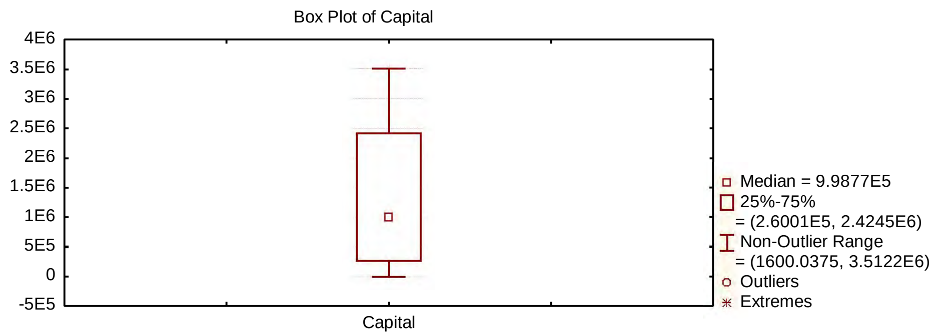


Figure 4.12: The Box and Whisker plot of market capital

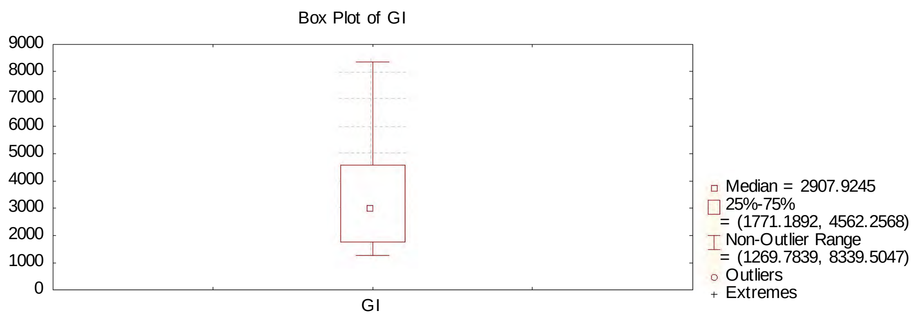


Figure 4.13: The Box and Whisker plot of general index

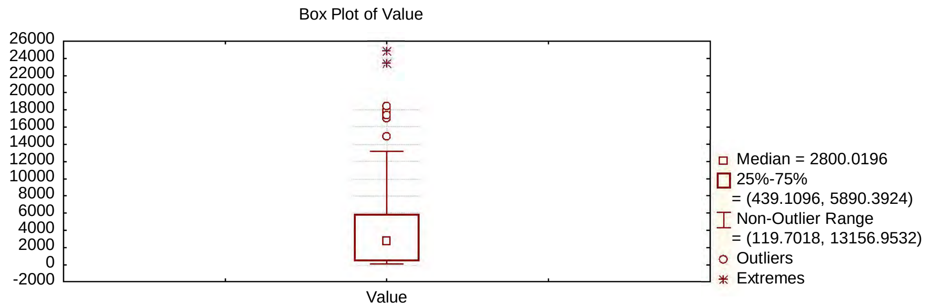


Figure 4.14: The Box and Whisker plot of value

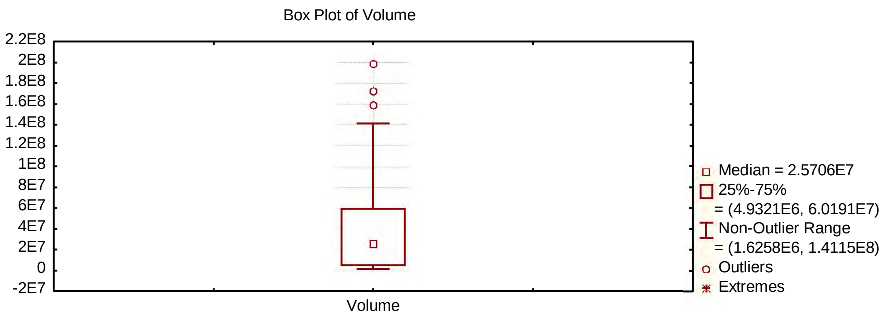


Figure 4.15: The Box and Whisker plot of volume

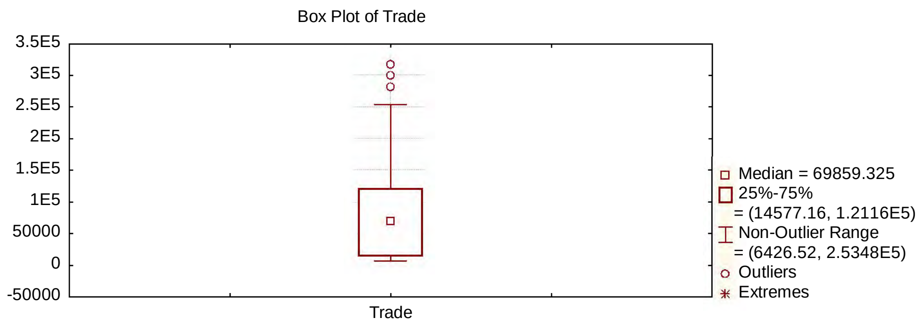


Figure 4.16: The Box and Whisker plot of trade

The Box and Whisker plot of market capital, general index, value, volume, and trade respectively (Figures 4.12, 4.13, 4.14, 4.15, and 4.16) reveal the essential statistics. Median capital is 9.9877E5, 25% to 75% frequency is in-between 2.6001E5 and 2.4245E6; non-outlier range is 1,600.0375 to 3.5122E6 of market capital; and it is not affected by outlier and extreme values. Median general index is 2,907.9245, 25% to 75% frequency is in-between 1,771.1892 and 4,562.2568, non-outlier range is 1,269.7839 to 8,339.5047; and it is not affected by outlier and extreme values also. Median market value is 2,800.0196, 25% to 75% frequency is in-between 439.1096 and 5,890.3924; non-outlier range is 119.7018 to 13,156.9532; and it is either affected by outlier and extreme values. Median market volume is 69,859.325, 25%-75% frequency is in-between 14,577.16 and 1.2116E5; non-outlier range is 6,426.52 to 2.5348E5; and it is either affected by outliers but not extreme values. Median market trade is 69,859.325, 25% to 75% frequency is in-between 14,577.16 and 1.2116E5; non-outlier range is 6,426.52 to 2.5348E5; and it is either affected by outliers but not extreme values.

4.4.1 Unit Root Test of Study Variables of DSE

To check the stationary of the series, unit root tests are conducted. ADF (Dickey and Fuller, 1979), PP test (Phillips and Perron, 1998), KPSS (Kwiatkowski et al., 1992), ERS (Elliott et al., 1996), and NP test (Ng. and Perron, 2001) are applied.

Table 4.9: Unit root test of study variables of DSE

Variables	**Deterministic terms**	**ADF (*P*-value)**	**PP (*P*-value)**	**KPSS [Critical value]***	**ERS [Critical value]***	**NP [Critical value]***
Market capital	Constant and linear trend	-1.59274 (0.79)	-1.74149 (0.73)	0.346798 [0.146]	19.41497 [5.642]	-4.71836 [-17.30]
Δ(Market capital)	Constant	-7.747261 (0.00)**	-11.1975 (0.00)**	0.11879 [0.463]	0.444132 [3.1154]	-56.7145 [-8.100]
General index	Constant and linear trend	-1.22071 (0.90)	-1.4494 (-3.451)	0.30754 [0.1460]	20.4768 [5.642]	-4.34904 [-17.30]
Δ(General index)	Constant	-7.78123 (0.00)**	-6.7700 (0.00)**	0.13408 [0.4630]	0.27728 [3.1154]	-85.4106 [-8.100]
Value	Constant and linear trend	-1.75874 (0.399)	-2.6536 (0.258)	0.3679 [0.146]	11.81244 [5.642]	-7.60684 [-17.30]
Δ(Value)	Constant	-10.74108 (0.00)**	-10.474 (0.00)**	0.034918 [0.4630]	0.199769 [3.115]	-121.667 [-8.100]
Volume	Constant and linear trend	-5.4066 (0.0001)**	-6.2602 (0.00)**	0.21027 [0.146]	1.79308 [5.642]	-46.0939 [-17.30]
Trade	Constant and linear trend	-2.235027 (0.465)	-3.4555 (0.049)**	0.40364 [0.146]	9.03074 [5.642]	-9.96441 [-17.30]
Δ(Trade)	Constant	-11.67007 (0.00)**	-11.584 (0.00)**	0.0242 [0.4630]	0.19191 [3.115]	-127.985 [-8.100]

Notes. [ ]* indicates the critical value at 5% level of significance and ( )** indicates the *P*-value at 5% level of significance of the respective test statistics. Δ represents first order difference.

Table 4.9 represents the unit root test of market capital, general index, value, volume, and trade of DSE. ADF, PP, KPSS, ERS, and NP tests results indicate that all variables are non-stationary by not rejecting the null hypothesis of unit-root at 5% levels of significance and critical values, but they are all stationary after first order differencing except volume data of DSE which is normally stationary. Therefore, first order differenced series is used for all variables except volume series in this analysis.

4.4.2 Empirical Results and Diagnostics

This section aims at determining the true lag order of the VAR model. Lutkepohl (1991) stated that a lag length, higher than the true lag length, increases the mean square forecast errors of the model. On the other hand, a lower order lag length than the true lag length generally causes autocorrelated errors. That is why the accuracy of forecasting of VAR models is very vital to detect the true lag length. Selection of true lag length needs calculation of several statistical criteria. We identify a VAR(p) model for the analysis by using penalty selection criteria, such as AIC and BIC. This analysis reveals the minimum value of AIC and BIC has got at the lag length of order two than that of any other lag lengths of orders. After that a VAR(2) model is identified, then the model estimation process is conducted. Since all the variables are stationary after first order difference except volume data series. In this type of situation, the researchers proposed different opinions. Sophia (2016) and Michael (1994) suggested considering the first difference of the variables if they are not cointegrated. Ömer (2016) proposed that there is no need to calculate differenced versions of all the variables, only transform the variables which are not stationary in level like $I(1)$. The Johansen cointegration test result among market capital, market volume, market value, trade, and DGI is presented in Table 4.10. From Table 4.10, we observe that the trace statistics are greater than 5% critical value and statistically significant at 5% level. Therefore, we may reject the null hypothesis that none of the series and at most one of the series are cointegrated. Since none of the series are cointegrated, we estimate the VAR models using first order differenced series of all variables except volume data series only. The model estimation results from the VAR(2) model are given in Table 4.11.

Table 4.10: Johansen cointegration test results of the variables

Sample (adjusted): 2004:09 2013:07
Included observations: 107 after adjustments
Trend assumption: Linear deterministic trend
Series: CAPITAL, VOLUME, VALUE, TRADE, and DGI
Lags interval (in first differences): 1 to 2

Unrestricted Cointegration Rank Test (Trace)

Hypothesized No. of CE(s)	Eigenvalue	Trace Statistic	0.05 Critical Value	Prob.**
None *	0.390746	105.6329	69.81889	0.0000
At most one *	0.259680	52.61223	47.85613	0.0167

* denotes rejection of the hypothesis at the 0.05 level
**MacKinnon-Haug-Michelis p-values

Table 4.11: Model estimation results from VAR(2) model

Variables	DCAPITAL	DGI	DTRADE	DVALUE	VOLUME
DCAPITAL(-1)	-0.570857	5.76E-06	0.009318	0.000680	2.851956
SE	(0.11300)	(0.00019)	(0.02884)	(0.00202)	(21.5178)
t-statistics	[-5.05199]*	[0.02973]*	[0.32309]*	[0.33700]	[0.13254]
DCAPITAL(-2)	-0.227448	-2.12E-05	0.012570	0.001288	1.128641
SE	(0.11183)	(0.00019)	(0.02854)	(0.00200)	(21.2953)
t-statistics	[-2.03391]	[-0.11064]*	[0.44038]*	[0.64548]*	[0.05300]
DGI(-1)	290.6657	0.370781	-17.28334	-1.020391	-7,478.008
SE	(68.9188)	(0.11814)	(17.5910)	(1.23011)	(13,124.2)
t-statistics	[4.21751]	[3.13860]	[-0.98251]	[-0.82951]	[-0.56979]
DGI(-2)	32.37685	-0.136865	-10.16488	-0.536241	-15,827.73
SE	(65.7113)	(0.11264)	(16.7723)	(1.17286)	(12,513.4)
t-statistics	[0.49271]	[-1.21509]*	[-0.60605]	[-0.45721]	[-1.26487]
DTRADE(-1)	-4.342340	0.001977	-0.752422	-0.060810	-204.8256
SE	(2.48029)	(0.00425)	(0.63308)	(0.04427)	(472.319)
t-statistics	[-1.75074]	[0.46500]*	[-1.18852]	[-1.37362]*	[-0.43366]
DTRADE(-2)	-1.335990	-0.001847	-0.044877	0.012303	-247.7935
SE	(1.50141)	(0.00257)	(0.38322)	(0.02680)	(285.912)
t-statistics	[-0.88982]	[-0.71773]*	[-0.11710]	[0.45911]*	[-0.86668]
DVALUE(-1)	64.86912	0.045059	7.036459	0.733140	619.5228
SE	(27.5907)	(0.04729)	(7.04231)	(0.49246)	(5,254.07)
t-statistics	[2.35113]	[0.95275]*	[0.99917]	[1.48874]	[0.11791]
DVALUE(-2)	27.31420	0.059777	-4.865443	-0.558370	769.0154
SE	(23.0874)	(0.03957)	(5.89289)	(0.41208)	(4,396.52)
t-statistics	[1.18308]	[1.51049]*	[-0.82565]	[-1.35501]	[0.17491]
VOLUME(-1)	0.001878	-2.76E-06	3.73E-05	3.25E-06	1.010229
SE	(0.00138)	(2.4E-06)	(0.00035)	(2.5E-05)	(0.26289)
t-statistics	[1.36047]*	[-1.16631]*	[0.10592]*	[0.13184]*	[3.84280]
VOLUME(-2)	-0.001963	2.25E-06	-0.000272	-1.88E-05	-0.168293
SE	(0.00144)	(2.5E-06)	(0.00037)	(2.6E-05)	(0.27344)
t-statistics	[-1.36716]*	[0.91466]*	[-0.74192]*	[-0.73336]*	[-0.61548]*

Variables	DCAPITAL	DGI	DTRADE	DVALUE	VOLUME
Constant	33721.33	39.46907	10,961.83	683.5573	7,978,627
SE	(18,195.6)	(31.1896)	(4,644.29)	(324.766)	(3,464,975)
t-statistics	[1.85327]	[1.26546]	[2.36028]	[2.10477]	[2.30265]
AIC	26.46909	13.73139	23.73801	18.41743	36.96765
BIC	26.74387	14.00617	24.01278	18.69221	37.24243

Notes. Sample (adjusted): 2004:09 2013:07, included observations: 107 after adjusting endpoints; standard errors in ( ) and *t*-statistics in [ ] and [ ]* indicates that the estimated coefficients are statistically significant at 5% level of significance. DCAPITAL, DGI, DTRADE, and DVALUE represent the first order differenced series from the respective original time series.

After the estimation of a suitable VAR (2) model, the diagnostic checking is conducted. Several methods control the robustness of the model and graphical analysis tools and statistical tests of the residuals used for diagnostic checking. Table 4.12 reveals the results of normality (H_0: residuals are multivariate normal) and Table 4.13 shows heteroscedasticity tests of the residuals. Table 4.14 and Figure 4.17 show the root of the characteristic polynomial of the estimated VAR model which confirms the stability condition. Figure 4.17 shows the correlations of the estimated residuals of the VAR (2) model.

Table 4.12: Normality test of the estimated residuals of VAR (2) model

Component	Skewness	Chi-square	df	Probability
1	-2.022983	72.98220	1	0.0000
2	0.229877	0.942372*	1	0.3317
3	-0.214868	0.823334*	1	0.3642
4	-0.244362	1.064879*	1	0.3021
5	0.613907	6.721059*	1	0.0095
Joint		82.53385	5	0.0000
Component	**Kurtosis**	**Chi-square**	**df**	**Probability**
1	14.72666	613.0858	1	0.0000
2	4.272266	7.216529	1	0.0072
3	5.947517	38.73337	1	0.0000
4	5.232218	22.21498	1	0.0000
5	3.460395	0.945004*	1	0.3310
Joint		682.1957	5	0.0000
Component	**Jarque-Bera**	**df**	**Probability**	
1	686.0680	2	0.0000	
2	8.158901	2	0.0169	
3	39.55670	2	0.0000	
4	23.27986	2	0.0000	
5	7.666063	2	0.0216	
Joint	764.7295	10	0.0000	

Note. VAR residual normality tests [Cholesky (Lutkepohl)].

From Table 4.12, we observe that the estimated residuals of the VAR(2) model were lack of multivariate normal distribution and statistically significant at the 5% level of significance except (*) marked statistics. Startz (2009) proposed that the non-normality of residuals of a VAR model is not very important. He argued that a large sample can be highly significant of Jarque-Bera statistic, even though the estimated VAR residuals are not very far from normality.

Table 4.13: VAR residual heteroscedasticity tests

Joint Test		
Chi-square	**df**	**Probability**
486.1293*	300	0.0000

Note. VAR residual heteroscedasticity tests: no cross-terms (only levels and square).

Table 4.13 indicates that residuals from the VAR(2) model reject the null hypothesis of no ARCH effects (heteroscedasticity problem) at 1% level (Chi-sq = 486.1293, p. value < 0.01).

Table 4.14: Stability test of roots of characteristic polynomial of estimated VAR model

Root	Modulus
0.925801	0.925801
0.110128 – 0.628654i	0.638227
0.110128 + 0.628654i	0.638227
-0.081331 – 0.488793i	0.495513
-0.081331 + 0.488793i	0.495513
-0.288446 + 0.391117i	0.485977
-0.288446 – 0.391117i	0.485977
0.194636 – 0.325181i	0.378980
0.194636 + 0.325181i	0.378980
-0.004904	0.004904

Notes. Endogenous variables: D (capital), D (GI), D (trade), D (value), and volume; D represents I(1); exogenous variables: constant; and lag specification: 1and 2.

Table 4.14 and Figure 4.17 represent that no root lies outside the unit circle. Therefore, VAR(2) model satisfies the stability condition.

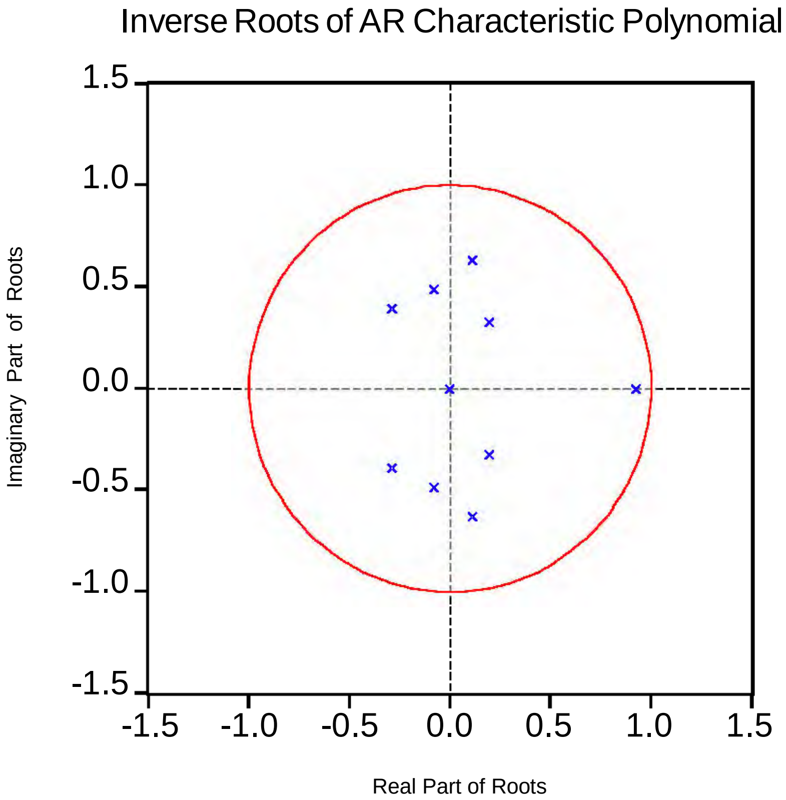


Figure 4.17: Inverse roots of AR characteristic polynomial of the estimated VAR(2) model

The correlations of the estimated residuals of the VAR(2) model are shown in Figure 4.18. As we see, most of the spikes from the estimated residuals of the VAR(2) model fall within the three-sigma confidence interval. Therefore, it may be free from outliers and extreme values. To see the dynamics of the variables, impulse response analysis and Granger causality tests are applied. Figure 4.19 shows the combined graph of the impulse responses of each variable of the estimated VAR(2) model. We observe that from the graph, stock capital has an immediate effect on the general index, trade, current value, and volume. Similarly, general index, trade, current value, current volume, and stock capital have an immediate effect on all others except volume series of DSE. Stock volume has only a direct impact on the general index of DSE. The Granger causality test of each variable of DSE is presented in Table 4.15.

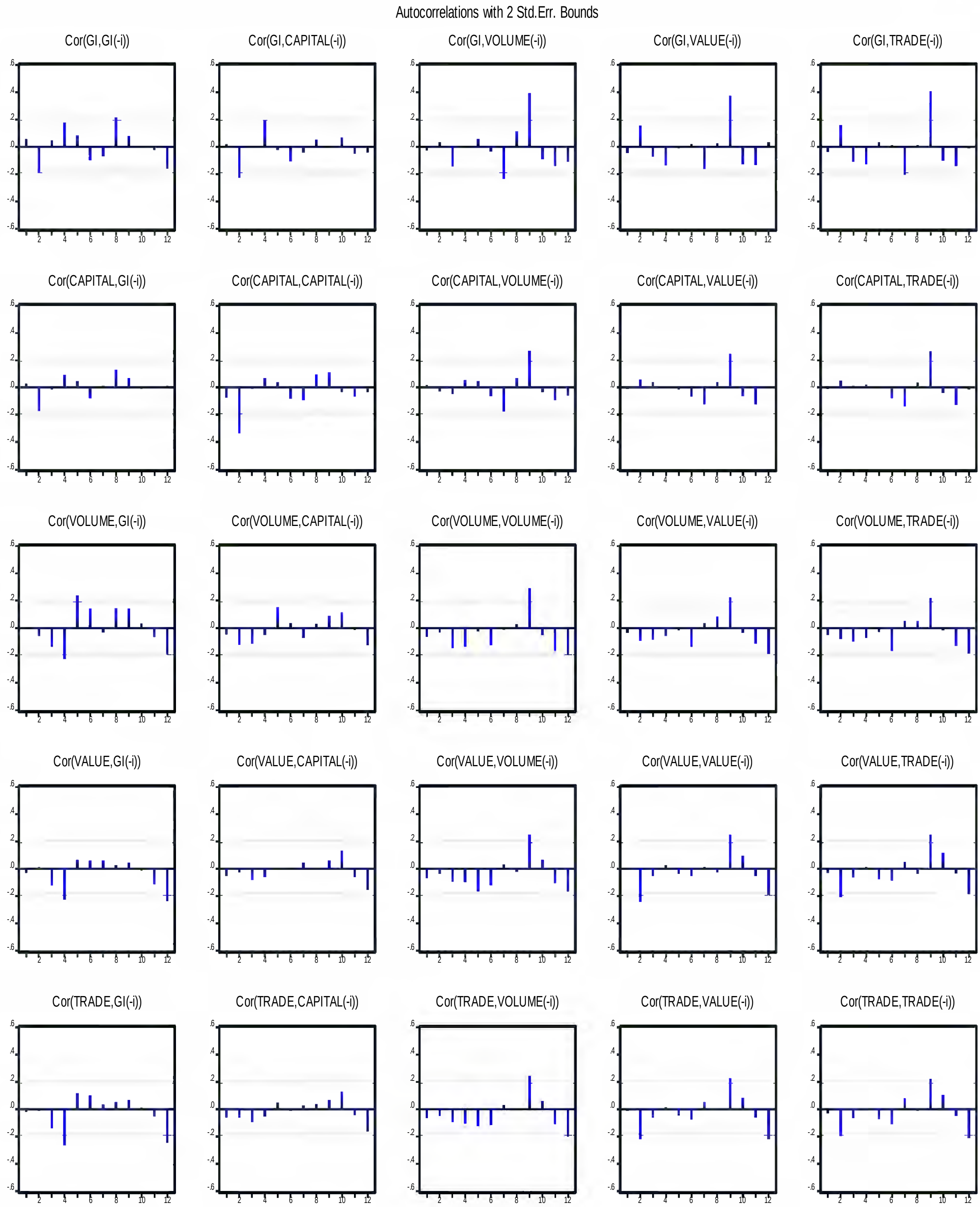


Figure 4.18: Correlations of the estimated residuals of VAR(2) model

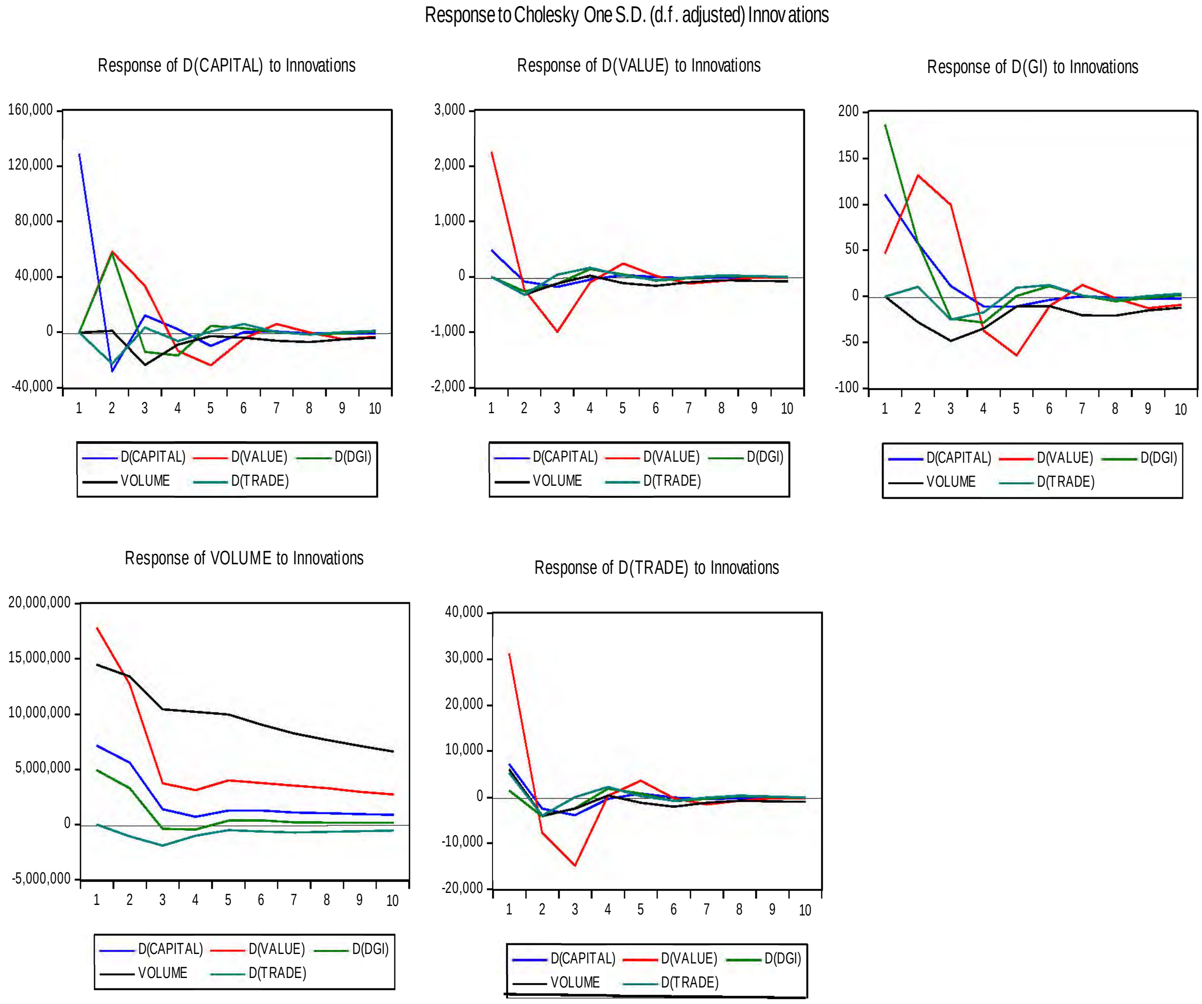


Figure 4.19: The combined graph of the impulse responses of the estimated VAR(2) model

Table 4.15: Pairwise Granger Causality tests

Null Hypothesis	Obs	*F*-Statistic	Probability
DGI does not Granger Cause CAPITAL	108	14.4312*	3.0E-06
CAPITAL does not Granger Cause GI		0.99219	0.37428
TRADE does not Granger Cause CAPITAL	108	19.2568*	7.9E-08
CAPITAL does not Granger Cause TRADE		3.90001	0.02330
VALUE does not Granger Cause CAPITAL	108	17.3154*	3.3E-07
CAPITAL does not Granger Cause VALUE		1.13685	0.32482
VOLUME does not Granger Cause CAPITAL	108	5.40554*	0.00586
CAPITAL does not Granger Cause VOLUME		19.6973*	5.7E-08
TRADE does not Granger Cause GI	108	31.3317*	2.3E-11
DGI does not Granger Cause TRADE		4.80681*	0.01010
VALUE does not Granger Cause GI	108	32.6324*	1.1E-11
DGI does not Granger Cause VALUE		2.27186	0.10826
VOLUME does not Granger Cause GI	108	6.17087*	0.00294
DGI does not Granger Cause VOLUME		9.15604*	0.00022

Null Hypothesis	Obs	*F*-Statistic	Probability
VALUE does not Granger Cause TRADE	108	0.39541	0.67442
TRADE does not Granger Cause VALUE		0.17806	0.83715
VOLUME does not Granger Cause TRADE	108	2.51300	0.08598
TRADE does not Granger Cause VOLUME		0.76554	0.46771
VOLUME does not Granger Cause VALUE	108	1.91118	0.15311
VALUE does not Granger Cause VOLUME		0.52185	0.59498

Note. Lags: 2 and (*) marked that *F*-Statistics are statistically significant at the 5% level of significance.

The test results indicate that there are bivariate causal relationships among the variables marked as (*) by rejecting the null hypothesis of no Granger causality. Data from June 2004 to June 2012 are used for training samples and from July 2012 to July 2013 are used for testing samples and compared the results of the VAR(2) model with the univariate auto ARIMA models. The auto ARIMA models of capital, DGI, value, volume, and trade data series are estimated using auto.arima( ) function of R Package 'forecast', Version 8.13 (Rob J., Hyndman et al., 2020). Although volume series is stationary at level, auto ARIMA selected ARIMA(1,1,1) models for capital, DGI, value, volume, and trade data series of DSE, respectively. Auto ARIMA estimation results of capital, DGI, value, volume and trade data series of DSE are presented in Table 4.16.

Table 4.16: Auto ARIMA estimation results of capital, DGI, value, volume, and trade data series of DSE

Method: Least Squares Sample (adjusted): 2004M08 to 2013M07

ARIMA (1, 1, 1) model of capital

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	22688.81	13822.48	1.641443	0.1037
AR(1)	0.159395	1.037946	0.153568	0.8782
MA(1)	-0.247618	1.019548	-0.242871	0.8086

ARIMA (1, 1, 1) model of DGI

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	29.77008	40.38824	0.737098	0.4627
AR(1)	0.041349	0.157112	0.263182	0.7929
MA(1)	0.597280**	0.126685	4.714672	0.0000

ARIMA (1, 1, 1) model of market value

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	52.39103	142.6687	0.367222	0.7142

Variable	Coefficient	Std. Error	t-Statistic	Prob.
AR(1)	0.552505*	0.264047	2.092453	0.0388
MA(1)	-0.736540**	0.214556	-3.432856	0.0009

ARIMA (1, 1, 1) model of volume

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	1048366**.	129931.5	8.068603	0.0000
AR(1)	0.466581**	0.086773	5.377020	0.0000
MA(1)	-0.999806**	0.001128	-886.3726	0.0000

ARIMA (1, 1, 1) model of trade

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	1046.692	1309.025	0.799597	0.4257
AR(1)	0.406864*	0.168482	2.414887	0.0175
MA(1)	-0.771144**	0.117526	-6.561491	0.0000

Note. ** and * indicate that the estimated coefficients are statistically significant at 1% and 5% levels of significance.

After the estimation of auto ARIMA models, the forecasting performances are checked using the RMSE metrics. RMSE statistics for overall samples (June 2004 to July 2013) of VAR(2) and auto ARIMA model of capital, DGI, value, and trade data series of DSE are shown in Table 4.17.

Table 4.17: RMSE statistics for overall samples of VAR(2) and auto ARIMA models

Variable	VAR(2)	ARIMA (1, 1, 1)
Δ(Capital)	34.71275	38.80718
Δ(DGI)	1.437179	1.53407
Δ(Value)	17.53741	4.769211
Volume	4.637578	523.2658
Δ(Trade)	479.0224	18.09644

Note. Overall samples (Training & testing): June 2004 to July 2013

Table 4.17 shows that RMSE statistics for VAR(2) models for market capital, DGI, and volume data series are minimal compared to ARIMA(1,1,1) models. Accordingly, the forecasting performance of the VAR(2) model is quite better than that of ARIMA(1,1,1). But, the error distribution of the estimated residuals is lacking normal distribution and ARCH effect problems are found there. Unfortunately, the DGI count was suspended after July 31, 2013. So, the forecasting of DGI and its associated variables are less valuable for the near future analysis of DSE portfolios. So, in the next section, univariate ARIMA with GARCH (ARCH) family models are estimated with the newest indicators of DSE like DSEX, DSES, and DSE30 indices for forecasting purposes.

4.5 Univariate Modeling and Forecasting

In this study, an attempt is made to reveal the usefulness of univariate time series analysis as both an analytical and forecasting tool for DSEX, DSES, and DSE30 indices time series. The data set covers the monthly DSEX, DSES, and DSE30 indices from January 2014 to December 2018 (N.B: Holiday effect consists of every week). To check the stationary condition of DSEX, DSES, and DSE30 indices, Kwiatkowski Phillips Schmidt Shin (KPSS) test is applied (Kwiatowski et al., 1992). If the KPSS (LM) statistic is greater than the critical value for alpha levels of 5% then the null hypothesis is rejected; the series is non-stationary. The KPSS test of DSEX, DSES, and DSE30 indices is presented in Table 4.18. From Table 4.18, KPSS tests suggest that DSEX, DSES, and DSE30 indices are stationary at the 5% level.

Table 4.18: KPSS test of DSEX, DSES, and DSE30 indices

Variables	Deterministic Terms	KPSS Test (LM) Statistics	Asymptotic critical values at 5%	Remarks
DSEX	Constant	0.628	0.463	Stationary
DSES	Constant	0.729	0.463	Stationary
DSE30	Constant	0.612	0.463	Stationary

To forecast DSEX, DSES, and DSE30 indices, ARIMA, ARIMA with the GARCH family, ANN, and SVM models are estimated. Using ACF and PACF, proper ARIMA models are chosen. To choose the best fitted ARIMA Model, the minimum value of AIC and BIC is considered. Total suitable auto ARIMA model selection is conducted by using R Package ‘forecast’, Version 8.13 (Rob J. et al., 2020). The relevant R codes are included in the annexure. ARIMA with the GARCH family models are conducted using EViews 7, ANN and SVM models are conducted using STATISTICA 12.0. Finally, reliable models are proposed for

DSEX, DSES, and DSE30 indices forecasting that generate minimum RMSE compared to other models.

4.5.1 Modeling and Forecasting of DSEX

ARIMA, ARIMA with the GARCH family, ANN and SVM models of DSEX index are trained during the period January 2014 to December 2017, and ARIMA, ARIMA with the GARCH family, ANN and SVM models are tested during the period January 2018 to December 2018.

4.5.1.1 ARIMA Model of DSEX

The best fitted ARIMA model is estimated by using auto.arima( ) function of R Package 'forecast', Version 8.13 (Rob J. et al., 2020). Although DSEX is stationary at the 5% level of significance, auto.arima( ) function of R Package 'forecast' selects ARIMA(1,1,0) model. So, the ARIMA(1,1,0) model of DSEX is estimated. The summary of the ARIMA(1,1,0) model of DSEX during the training period is presented in Table 4.19. The coefficient AR(1) is approximately significant at the 5% level and the Durbin-Watson stat is approximately close to 2. So, the estimated residuals from ARIMA(1,1,0) models are not autocorrelated. The RMSE of the test period is much greater than the RMSE of the training period. The actual, fitted, and residual plot of the ARIMA(1,1,0) model of DSEX during training is shown in Figure 4.20. From Figure 4.20, we observe that there are huge differences between actual and fitted data points. So, out-of-sample forecasts from the ARIMA(1,1,0) model will not be suitable.

Table 4.19: Summary of ARIMA(1,1,0) model of DSEX during training period

Variable	Coefficient	Std. Error	t-Statistic	Probability
ϕ_1	0.280959	0.139876	2.008626	0.050
AIC	13.00851			
BIC	13.04827			
Durbin-Watson stat	1.870283			
RMSE	158.1555			
RMSE (Test)	720.4677			

Note. Dependent Variable: D(DSEX), Method: Least Squares, Sample (training): 2014:01 2017:12, Sample (testing): 2018:01 2018:12

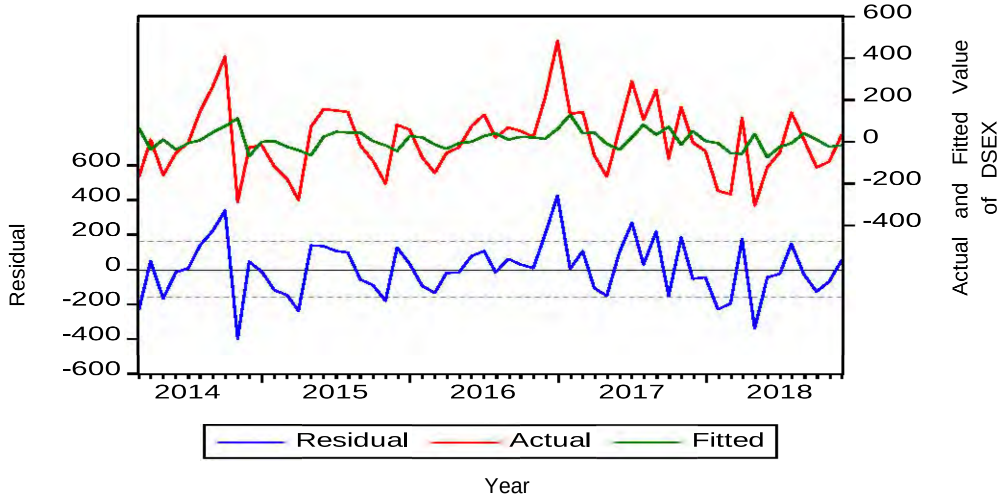


Figure 4.20: Actual, fitted, and residual plot of ARIMA(1,1,0) model of DSEX

4.5.1.2 ARIMA with GARCH Family Model of DSEX

Various types of GARCH family models are estimated using the ML-ARCH (Marquardt) method. Minimum AIC and BIC values are used to select the best performing GARCH family models. The summary of the GARCH family models of DSEX is shown in Table 4.20. The lowest AIC and BIC value is found for the model EGARCH(1,1,2). Thus, the EGARCH(1,1,2) model is chosen from the GARCH family for the DSEX index. Since DSEX is stationary at level and ARIMA(1,1,0) does not fit well. As such, the finite mixture ARIMA(1,0,0) with EGARCH(1,1,2) model is chosen and the results are computed. Model summary for ARIMA(1,0,0) with EGARCH(1,1,2) for DSEX is presented in Table 4.21.

Table 4.20: Summary of GARCH family models of DSEX

Models	Coefficients with constant (Probability)			AIC	BIC
ARCH(1)	18241.17 (0.0085)	0.355 (0.166)		13.064	13.134
ARCH(2)	18759.26* (0.0384)	0.351281 (0.1861)	-0.01709 (0.9258)	13.097	13.203

Models	Coefficients with constant (Probability)						AIC	BIC
ARCH(3)	19869.57*	0.406115	-0.05837	-0.08973			13.081	13.222
	(0.0114)	(0.1481)	(0.4678)	(0.5015)				
GARCH(1,1)	18986.35	0.354523	-0.02624				13.097	13.203
	(0.2754)	(0.1803)	(0.9614)					
GARCH(2,1)	24299.39	0.48735	-0.0394	-0.23094			13.079	13.220
	(0.2034)	(0.1032)	(0.9013)	(0.2818)				
GARCH(2,2)	18401.7	0.273167	-0.20054	0.527794	-0.27156		13.094	13.270
	(0.154)	0.3374	(0.000)	0.3356	(0.4677)			
TARCH(1,1,0)	18363.93*	0.674796	-0.71158				13.037	13.143
	(0.009)	(0.147)	(0.107)					
TARCH(1,1,1)	19007.9*	0.5777	-0.0775	-0.6411			13.044	13.185
	(0.0105)	(0.1372)	(0.4672)	(0.0816)				
EGARCH(1,1,1)	9.4929	0.4481	0.3799	0.0222			13.067	13.208
	(0.0895)	(0.2332)	(0.0800)	(0.9675)				
EGARCH(1,1,2)	**3.1680***	**-0.2406***	**0.3425***	**-0.6608***	**-0.1239**	**0.7559***	**12.957**	**13.168**
	(0.000)	**(0.000)**	**0.0003**	**0.0158**	**0.5384**	**(0.000)**		

Note. Dependent Variable: D(DSEX), Method: ML - ARCH (Marquardt), Sample (training): 2014:01 2018:12. * indicates significant coefficient at the 5% level.

Table 4.21: Summary of ARIMA(1,0,0) with EGARCH(1,1,2) model of DSEX

	Coefficient	Std. Error	z-Statistic	Probability
ARIMA(1,0,0) Equation				
C	5015.148*	515.1583	9.735157	0.0000
ϕ_1	0.959186*	0.034760	27.59430	0.0000
EGARCH(1,1,2) Equation				
	Coefficient	**Std. Error**	**z-Statistic**	**Probability**
ω	4.411477*	0.150011	29.40776	0.0000
β_1	0.142060	0.483483	0.293827	0.0689

	Coefficient	Std. Error	z-Statistic	Probability
α_1	0.413625*	0.255850	1.616675	0.0059
β_2	-0.362095*	0.376184	-0.962546	0.0358
α_2	-0.182528*	0.280591	-0.650510	0.0154
γ	0.576789*	0.053638	10.75327	0.0000

Training sample		Test sample	
R-square	0.919	R-square	0.811
AIC	13.160	AIC	12.912
BIC	13.442	BIC	13.235
F-statistic	83.617*	F-statistic*	2.454
Prob(F-statistic)	0.000	Prob(F-statistic)	0.021
RMSE	162.509	RMSE	120.297

Note. Dependent Variable: DSEX, Method: ML - ARCH (Marquardt), Sample (training): 2014:01 2017:12, Sample (test): 2018:01 2018:12. * indicates significant coefficient at the 5% level.

According to Table 4.21, the R-square value of the training period is 0.919, whereas the R-square value of the testing period is 0.811. The higher value of R-square ensures that the model is performed well in the training and test period. The overall statistics of the training period and the testing period are impressively good. Therefore, ARIMA(1,0,0) with EGARCH(1,1,2) is one of the suitable models for out-of-sample forecasting of the DSEX index. Figure 4.21 illustrates the actual, fitted, and residual plots of ARIMA(1,0,0) with EGARCH(1,1,2) model of DSEX from January 2014 to December 2020.

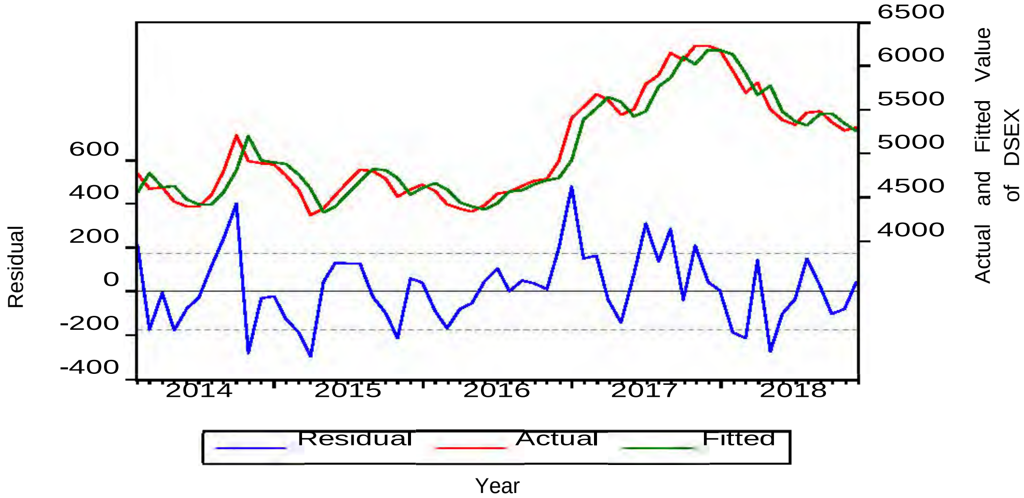


Figure 4.21: Actual, fitted, and residual plot of ARIMA(1,0,0) with EGARCH(1,1,2) model of DSEX

According to Figure 4.21, the actual, fitted, and residual plots of ARIMA(1,0,0) with EGARCH(1,1,2) model of DSEX indicate that the actual and fitted data are almost identical. Hence, the fitting of the model is reasonable. Figures 4.22 and Figure 4.23 show the estimated residual and standardized residual from the ARIMA(1,0,0) with EGARCH(1,1,2) model of DSEX. Both plots suggest that the distribution is close to symmetrical. When the p-value is less than or equal to 0.05, the Jarque-Bera test rejects the hypothesis of normality. According to Figure 4.22, the Jarque-Bera test statistic value is 0.6294 while the p-value is 0.720. In this case, the p-value is above 0.05, which indicates that the estimated residuals from the ARIMA(1,0,0) with GARCH(1,1,2) model of DSEX do not reject the hypothesis of normality. Hence, the error distribution of ARIMA(1,0,0) with EGARCH(1,1,2) model is normal.

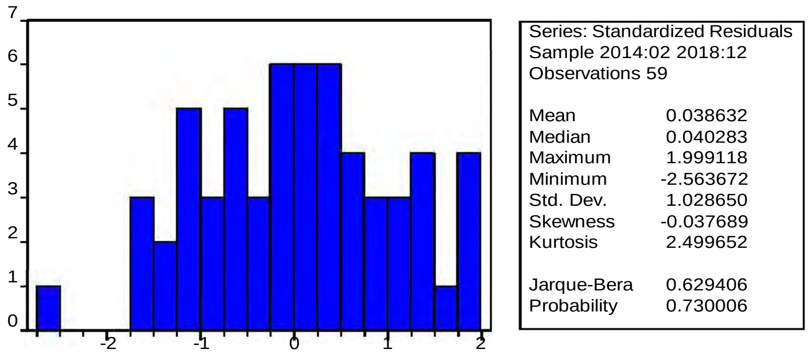


Figure 4.22: Histogram of residuals of ARIMA(1,0,0) with EGARCH(1,1,2) model of DSEX

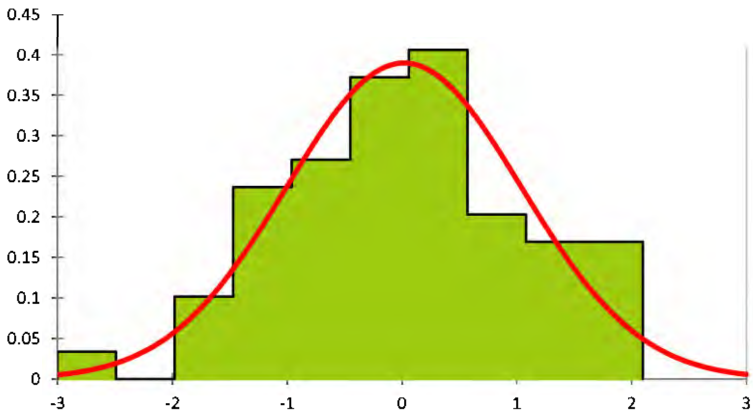


Figure 4.23: Histogram of the standardized residuals of ARIMA(1,0,0) with EGARCH(1,1,2) model of DSEX

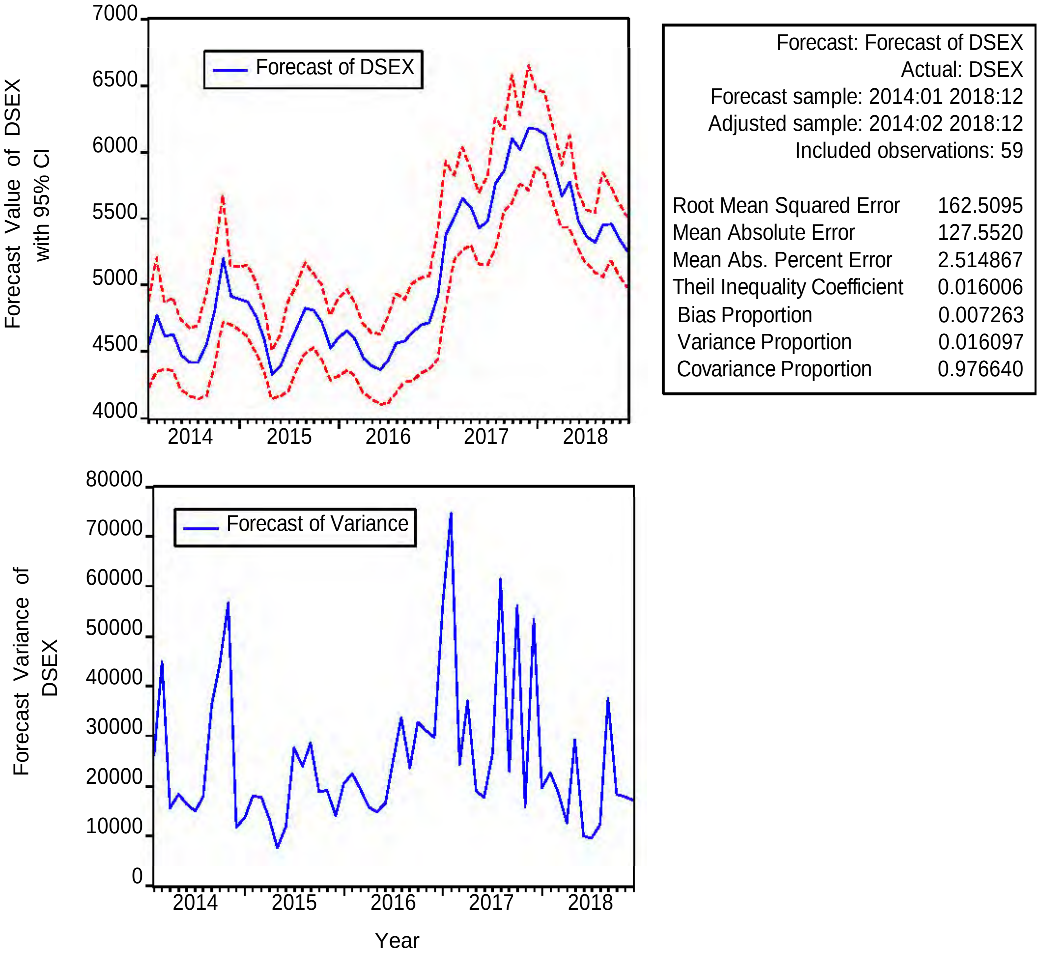


Figure 4.24: Static forecasting performance of ARIMA(1,0,0) with EGARCH(1,1,2) model of DSEX

Figure 4.24 illustrates the static forecasting performance of ARIMA(1,0,0) with EGARCH(1,1,2) model for DSEX index. ARIMA(1,0,0) with EGARCH(1,1,2) model has an RMSE of 162.50, which is comparatively lower than the RMSE of ARIMA and the GARCH family models of DSEX. The Theil inequality coefficient, biased proportion, and variance proportion are approximately close to zero. Therefore, the forecasting performance of ARIMA(1,0,0) with EGARCH(1,1,2) model of DSEX is quite reasonable. The out-of-sample forecast of ARIMA(1,0,0) with EGARCH(1,1,2) model of DSEX from January 2019 to December 2025 is shown in Table 4.22. Static forecasting is conducted for one-step forecast and dynamic forecasting is conducted for the multi-steps forecast.

Table 4.22: Out-of-sample forecast of ARIMA(1,0,0) with EGARCH(1,1,2) model of DSEX

Month	Dynamic Forecast	Static Forecast	Month	Dynamic Forecast	Static Forecast
Jan-19	4975.7207	5282.9058	Apr-22	5297.5773	5491.0691
Feb-19	4977.3299	5271.9775	May-22	5319.5473	5496.8935
Mar-19	4978.8734	5261.4951	Jun-22	5341.4799	5502.6489
Apr-19	4980.3539	5251.4406	Jul-22	5363.3754	5508.3368
May-19	4981.7740	5241.7964	Aug-22	5385.2342	5513.9590
Jun-19	4983.1361	5232.5459	Sep-22	5407.0564	5519.5169
Jul-19	4984.4426	5223.6729	Oct-22	5428.8425	5525.0121
Aug-19	4985.6958	5215.1621	Nov-22	5450.5925	5530.4461
Sep-19	4986.8979	5206.9986	Dec-22	5472.3069	5535.8203
Oct-19	4988.0509	5199.1684	Jan-23	5493.9858	5541.1360
Nov-19	4989.1568	5191.6577	Feb-23	5515.6296	5546.3946
Dec-19	4990.2176	5184.4535	Mar-23	5537.2386	5551.5972
Jan-20	4991.2351	5177.5434	Apr-23	5558.8130	5556.7452
Feb-20	4992.2111	5170.9153	May-23	5580.3531	5561.8397
Mar-20	4993.1472	5164.5578	Jun-23	5601.8592	5566.8819
Apr-20	4994.0451	5158.4597	Jul-23	5623.3315	5571.8729
May-20	4994.9064	5152.6105	Aug-23	5644.7705	5576.8137
Jun-20	4995.7325	5147.0000	Sep-23	5666.1763	5581.7054
Jul-20	4996.5250	5141.6185	Oct-23	5687.5492	5586.5490
Aug-20	4997.2850	5136.4567	Nov-23	5708.8896	5591.3455
Sep-20	4998.0141	5131.5056	Dec-23	5730.1977	5596.0958
Oct-20	4998.7134	5126.7565	Jan-24	5751.4738	5600.8009
Nov-20	4999.3841	5122.2012	Feb-24	5772.7182	5605.4616
Dec-20	5000.0275	5117.8319	Mar-24	5793.9312	5610.0788
Jan-21	4963.3796	5394.0275	Apr-24	5815.1130	5614.6534
Feb-21	4985.9390	5401.1455	May-24	5836.2641	5619.1861
Mar-21	5008.4574	5408.1587	Jun-24	5857.3845	5623.6777
Apr-21	5030.9352	5415.0704	Jul-24	5878.4747	5628.1291
May-21	5053.3724	5421.8836	Aug-24	5899.5349	5632.5409
Jun-21	5075.7695	5428.6011	Sep-24	5920.5655	5636.9139
Jul-21	5098.1265	5435.2257	Oct-24	5941.5666	5641.2489
Aug-21	5120.4437	5441.7601	Nov-24	5962.5386	5645.5464
Sep-21	5142.7214	5448.2068	Dec-24	5983.4818	5649.8071
Oct-21	5164.9597	5454.5681	Jan-25	6004.3964	5654.0318
Nov-21	5187.1589	5460.8466	Feb-25	6025.2828	5658.2210
Dec-21	5209.3193	5467.0442	Mar-25	6046.1411	5662.3753
Jan-22	5231.4411	5473.1633	Apr-25	6066.9718	5666.4954
Feb-22	5253.5245	5479.2059	May-25	6087.7750	5670.5818
Mar-22	5275.5698	5485.1738	Jun-25	6108.5511	5674.6351

Month	Dynamic Forecast	Static Forecast	Month	Dynamic Forecast	Static Forecast
Jul-25	6129.3002	5678.6559	Oct-25	6191.3893	5690.5284
Aug-25	6150.0228	5682.6447	Nov-25	6212.0337	5694.4243
Sep-25	6170.7191	5686.6020	Dec-25	6232.6526	5698.2902

4.5.1.3 ANN Models of DSEX

The selection of input variables is an essential part of developing a reasonable ANN model. In this analysis, the lagged variables from DSEX, (x_{t-1}, x_{t-2}, ..., x_{t-p}) are considered as the input variables, where p is the time delay. The input variables for training and testing ANN models of the output variable DSEX is $x_t = f(x_{t-1}, x_{t-2}, \quad x_{t-3})$. Different ANN models are trained and tested using the software STATISTICA 12. Radial Basis Function (RBF) networks and Multilayer Perceptron (MLP) networks are applied. Different training algorithms like BFGS and RBFT are used. We have applied different types of activation functions like Gaussian, Exponential, Tanh, etc. which are treated as hidden activation and output activation. ANN model's summary of DSEX is presented in Table 4.23.

Table 4.23: ANN models summary of DSEX

Net Name	RBF 3-13-1	MLP 3-10-1	RBF 3-14-1
Training Performance	0.7235	0.7798	0.7199
Test Performance	0.7409	0.7056	0.7249
Overall Performance	0.7322	0.7427	0.7224
Training error	0.0143	0.0118	0.0145
Test error	0.0084	0.0092	0.0088
Training Algorithm	RBFT	BFGS 24	RBFT
Hidden Activation	Gaussian	Exponential	Gaussian
Output Activation	Identity	Tanh	Identity

Note. Output Variable: DSEX, Sample (training): 2014:01 2017:12, Sample (test): 2018:01 2018:12

As shown in Table 4.23, we observe that MLP nets perform better than RBF nets. The test error of the RBF net is less than that of the MLP net, but the training error of MLP net is less than that of RBF. Therefore, MLP 3-10-1 net is used to execute the out-of-sample forecasts. The out-of-sample forecasts of the ANN model of DSEX from January 2019 to December 2025 are presented in Table 4.24.

Table 4.24: Out-of-sample forecast of ANN model of DSEX

Month	Forecast of DSEX	Month	Forecast of DSEX
Jan-19	4299.26	Dec-21	5834.59
Feb-19	4427.41	Jan-22	5820.03
Mar-19	4224.26	Feb-22	5805.47
Apr-19	4788.53	Mar-22	5790.91
May-19	5242.64	Apr-22	5776.35
Jun-19	5560.00	May-22	5761.80
Jul-19	5666.57	Jun-22	5747.24
Aug-19	7710.52	Jul-22	5732.68
Sep-19	11417.46	Aug-22	5718.12
Oct-19	11379.55	Sep-22	5703.57
Nov-19	7212.91	Oct-22	5689.01
Dec-19	5048.64	Nov-22	5674.45
Jan-20	4664.48	Dec-22	5659.89
Feb-20	8215.97	Jan-23	5645.33
Mar-20	9907.76	Feb-23	5630.78
Apr-20	8575.33	Mar-23	5616.22
May-20	6472.63	Apr-23	5601.66
Jun-20	5803.27	May-23	5587.10
Jul-20	4285.49	Jun-23	5572.54
Aug-20	4051.57	Jul-23	5557.99
Sep-20	4338.20	Aug-23	5543.43
Oct-20	4525.01	Sep-23	5528.87
Nov-20	5358.41	Oct-23	5514.31
Dec-20	5064.83	Nov-23	5499.76
Jan-21	5994.72	Dec-23	5485.20
Feb-21	5980.16	Jan-24	5470.64
Mar-21	5965.61	Feb-24	5456.08
Apr-21	5951.05	Mar-24	5441.52
May-21	5936.49	Apr-24	5426.97
Jun-21	5921.93	May-24	5412.41
Jul-21	5907.38	Jun-24	5397.85
Aug-21	5892.82	Jul-24	5383.29
Sep-21	5878.26	Aug-24	5368.73
Oct-21	5863.70	Sep-24	5354.18
Nov-21	5849.14	Oct-24	5339.62

Month	Forecast of DSEX	Month	Forecast of DSEX
Nov-24	5325.06	Jun-25	5223.16
Dec-24	5310.50	Jul-25	5208.60
Jan-25	5295.95	Aug-25	5194.04
Feb-25	5281.39	Sep-25	5179.48
Mar-25	5266.83	Oct-25	5164.92
Apr-25	5252.27	Nov-25	5150.37
May-25	5237.71	Dec-25	5135.81

4.5.1.4 SVM Models of DSEX

There is no suitable theory to select the optimal number of input nodes of SVM models. The same input structures of DSEX index are used for training and testing purposes. In this part, the lagged variables from DSEX, (x_{t-1}, x_{t-2}, ..., x_{t-p}) are considered as the input variables, where p is the time delay. The input variables for training and testing SVM models of the output variable DSEX is $x_t = f(x_{t-1}, x_{t-2}, x_{t-3})$. The selection of models and their parameters plays a crucial role to perform SVM models. Samsudin et al. (2010) emphasized that the accuracy of the estimation of SVM models depends upon a well setting of parameters such as hyper-parameters c, ε, and the kernel parameters. This study also emphasizes on application of the RBF kernel for its better performance. In past studies, the researchers found that the RBF kernel showed higher accuracy than those of other kernels of SVM models for time series forecasting (Ding et al., 2008; Eslamian et al., 2008; Wang et al., 2009). The RBF kernel nonlinearly maps data points into a higher dimensional space that can process nonlinear problems with negligible complexities. Different SVM regression models are trained and tested using the software STATISTICA 12. SVM models summary of DSEX is presented in Table 4.25.

Table 4.25: SVM models summary of DSEX

SVM Type	SVM Parameter		Kernel Type	Kernel Parameter (γ)	No. of support vectors	RMSE		
	c	ε				Train	Test	Overall
Regression	10.0	0.10	RBF	0.333	31	1695.268	1450.267	1634.359
Regression	10.0	0.10	Linear	-	28	1707.357	1430.59	1639.061
Regression	10.0	0.10	Polynomial (degree=3)	0.333	31	2013.535	1649.776	1924.486
Regression	10.0	0.10	Sigmoid	0.333	32	1928.507	2826.338	2200.585

Note. Output Variable: DSEX, Sample (training): 2014:01 2017:12, Sample (test): 2018:01 2018:12

The kernel type RBF was found to have a minimum RMSE compared to other kernel types in Table 4.25. Therefore, a regression-based SVM model of kernel type RBF is more appropriate for out-of-sample forecasting. The forecasts of the out-of-sample DSEX index from January 2019 until December 2025 are shown in Table 4.26.

Table 4.26: Out-of-sample forecast of SVM model of DSEX

Month	Forecast of DSEX	Month	Forecast of DSEX
Jan-19	4380.48	Apr-20	7324.56
Feb-19	4883.15	May-20	4847.43
Mar-19	5380.70	Jun-20	5668.87
Apr-19	5607.94	Jul-20	4506.54
May-19	5783.02	Aug-20	3430.62
Jun-19	7641.09	Sep-20	4311.45
Jul-19	11772.36	Oct-20	4697.42
Aug-19	11636.37	Nov-20	5348.61
Sep-19	8436.05	Dec-20	5348.09
Oct-19	6727.80	Jan-21	5328.72
Nov-19	4675.35	Feb-21	5247.72
Dec-19	3473.69	Mar-21	5166.71
Jan-20	8118.80	Apr-21	5085.70
Feb-20	9603.47	May-21	5004.69
Mar-20	8587.82	Jun-21	4923.68

Month	Forecast of DSEX	Month	Forecast of DSEX
Jul-21	4842.68	Oct-23	2655.47
Aug-21	4761.67	Nov-23	2574.46
Sep-21	4680.66	Dec-23	2493.45
Oct-21	4599.65	Jan-24	2412.44
Nov-21	4518.65	Feb-24	2331.44
Dec-21	4437.64	Mar-24	2250.43
Jan-22	4356.63	Apr-24	2169.42
Feb-22	4275.62	May-24	2088.41
Mar-22	4194.61	Jun-24	2007.41
Apr-22	4113.61	Jul-24	1926.40
May-22	4032.60	Aug-24	1845.39
Jun-22	3951.59	Sep-24	1764.38
Jul-22	3870.58	Oct-24	1683.37
Aug-22	3789.58	Nov-24	1602.37
Sep-22	3708.57	Dec-24	1521.36
Oct-22	3627.56	Jan-25	1440.35
Nov-22	3546.55	Feb-25	1359.34
Dec-22	3465.54	Mar-25	1278.34
Jan-23	3384.54	Apr-25	1197.33
Feb-23	3303.53	May-25	1116.32
Mar-23	3222.52	Jun-25	1035.31
Apr-23	3141.51	Jul-25	954.30
May-23	3060.51	Aug-25	873.30
Jun-23	2979.50	Sep-25	792.29
Jul-23	2898.49	Oct-25	711.28
Aug-23	2817.48	Nov-25	630.27
Sep-23	2736.48	Dec-25	549.27

4.5.1.5 Comparison of Different Models of DSEX

Table 4.27 illustrates the RMSE for the best fitting ARIMA with the GARCH family, ANN, and SVM models of DSEX for the overall samples between January 2014 and December 2018.

Table 4.27: RMSE of the estimated models of DSEX

Model Name	RMSE
ARIMA with GARCH family	162.5095
ANN	1599.3621
SVM	1634.3590

The RMSE of ARIMA with the GARCH family model is lower than ANN and SVM models of DSEX as shown in Table 4.27. Therefore, ARIMA with GARCH family model is the most reliable model to forecast the DSEX index.

4.5.2 Modeling and Forecasting of DSES

ARIMA, ARIMA with GARCH family, ANN and SVM models of the DSES index are trained from January 2014 to December 2017, and ARIMA, ARIMA with GARCH family, ANN and SVM models are tested from January 2018 to December 2018.

4.5.2.1 ARIMA Model of DSES

The best-fitted ARIMA model is estimated using the auto.arima( ) function of R Package 'forecast', Version 8.13 (Rob J. et al., 2020). However, auto.arima( ) function selects the ARIMA(1,1,0) model despite DSES stationary at 5% significance level. As a result, an ARIMA(1,1,0) model of DSES is estimated. Table 4.28 summarizes the model summary of the ARIMA(1,1,0) model of DSES during training. A 5% level of AR(1) does not indicate significance, and the Durbin-Watson statistic is close to 2. Therefore, the estimated residuals from ARIMA(1,1,0) models are not auto-correlated. The RMSE of the test period is slightly lower than the RMSE of the training period. The actual, fitted, and residual plot of the ARIMA(1,1,0) model of DSES during training is shown in Figure 4.25. As can be seen from Figure 4.25, there are large differences between the actual and fitted data points. So, out-of-sample forecasts from the ARIMA(1,1,0) model will not be suitable.

Table 4.28: Summary of ARIMA(1,1,0) model of DSES during training period

Variable	Coefficient	Std. Error	t-Statistic	Prob.
c	8.412078	6.885546	1.221701	0.2283
ϕ_1	0.241683	0.146147	1.653697	0.1053
AIC	10.01455			
BIC	10.09405			
Durbin-Watson stat	1.953044			
R-square	0.058516			

RMSE	34.63469
RMSE (Test)	31.53877

Note. Dependent Variable: D(DSES), Method: LS, Sample (training): 2014:01 2017:12, Sample (testing): 2018:01 2018:12

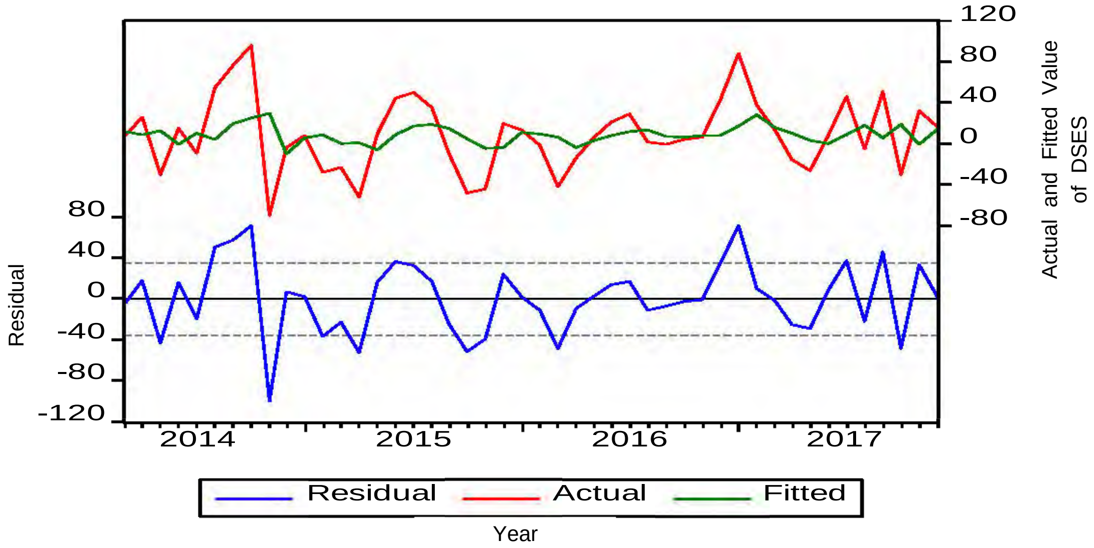


Figure 4.25: Actual, fitted, and residual plot of ARIMA(1,1,0) model of DSES

4.5.2.2 ARIMA with GARCH Family Model of DSES

Using ML - ARCH (Marquardt) method, various types of GARCH family models are estimated. To select the best performed GARCH family models, the minimum value of AIC and BIC is considered. The summary of the GARCH family models of DSES is exposed in Table 4.29. We observe that ARCH(2) has the lowest AIC and BIC values. For the DSES index, the ARCH(2) model is selected from the models of the GARCH family. Since DSES is stationary at level and ARIMA(1,1,0) model does not fit well. Therefore, the finite mixture of ARIMA(1,0,0) with ARCH(2) model is chosen and then estimated using the DSES index. The model summary of ARIMA(1,0,0) with the ARCH(2) model of DSES is presented in Table 4.30.

Table 4.29: Summary of GARCH family models of DSES

Models	Coefficients with constant (Prob.)					AIC	BIC
ARCH(1)	812.587*	0.358				9.989	10.157
	(0.007)	(0.184)					
ARCH(2)	**1015.216***	**0.386**	**-0.182**			**9.983**	**10.102**
	(0.003)	**(0.114)**	**(0.084)**				
ARCH(3)	1022.332*	0.326	-0.093	-0.056		10.008	10.165
	(0.016)	(0.210)	(0.625)	(0.789)			
GARCH(1,1)	1044.398	0.365	-0.204			10.024	10.142
	(0.190)	(0.159)	(0.703)				
GARCH(2,1)	1650.921	0.363	-0.354	-0.341		10.009	10.166
	(0.127)	(0.134)	(0.379)	(0.399)			
GARCH(2,2)	890.964	0.273	-0.206	0.202	-0.001	10.048	10.245
	(0.759)	(0.182)	(0.632)	(0.897)	(0.999)		
EGARCH(1,1,1)	5.067	0.356	0.374	0.227		10.033	10.190
	(0.164)	(0.411)	(0.189)	(0.656)			
EGARCH(2,1,1)	6.228	0.426	0.341	0.491	-0.441	10.038	10.235
	0.058	0.241	0.265	0.343	0.280		

Note. Dependent Variable: D(DSES), Method: ML - ARCH (Marquardt), Sample (training): 2014:01 2017:12. * indicates significant coefficient at the 5% level.

Table 4.30: Summary of ARIMA(1,0,0) with ARCH(2) model of DSES

ARIMA(1,0,0) Equation				
	Coefficient	**Std. Error**	**z-Statistic**	**Prob.**
C	1397.170*	348.7854	4.005815	0.0001
ϕ_1	0.962221*	0.059851	16.07698	0.0000
ARCH(2) Equation				
α_0	1647.742*	649.7655	2.535903	0.0112
α_1	0.206298	0.203872	1.011896	0.0516
α_2	-0.283332*	0.137789	-2.056265	0.0398
Training sample		**Test sample**		
R-square	0.893	R-square		0.831
AIC	10.001	AIC		10.324
BIC	10.197	BIC		10.526
F-statistic	87.859*	F-statistic		8.584*
Prob(F-statistic)	0.000	Prob(F-statistic)		0.007
RMSE	35.166	RMSE		25.722

Note. Dependent Variable: DSES, Method: ML - ARCH (Marquardt), Sample (training): 2014:01 2017:12, Sample (test): 2018:01 2018:12. * indicates significant coefficient at the 5% level.

Table 4.30 shows that the R-square values of the training and testing periods are respectively 0.893 and 0.831. Models performing well in training and testing periods are indicated by greater values of R-square. The overall model fitting statistics are significant at the 5% level for training and test samples. The overall statistics of the training period and the testing period are almost identical. For out-of-sample forecasting of the DSES index, ARIMA(1,0,0) with ARCH(2) is an appropriate model. The actual, fitted, and residual plot of ARIMA(1,0,0) with ARCH(2) model of DSES from January 2014 to December 2018 is presented in Figure 4.26.

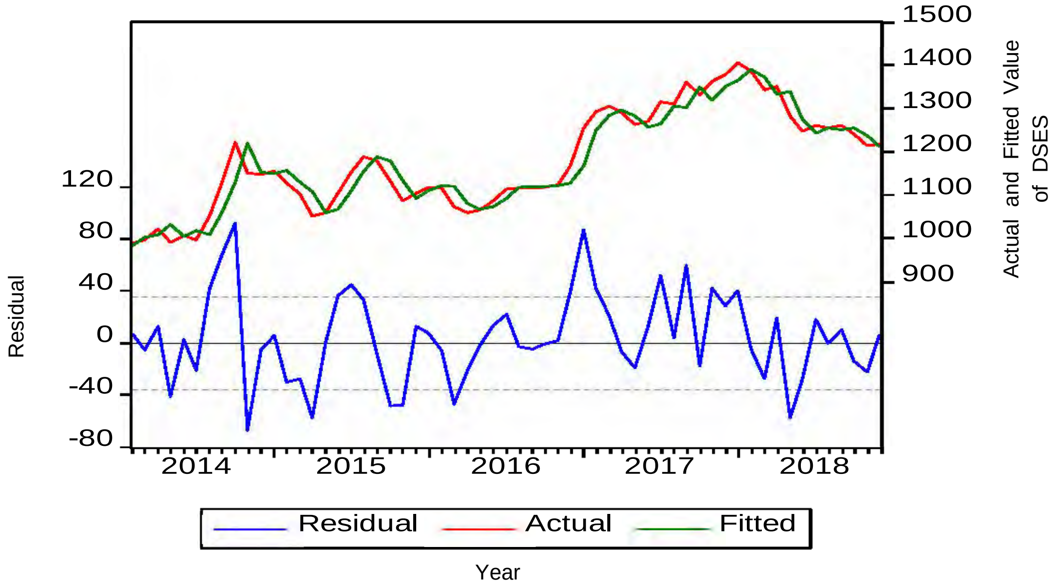


Figure 4.26: Actual, fitted, and residual plot of ARIMA(1,0,0) with ARCH(2) model of DSES

In figure 4.26, the actual, fitted, and residual plots of the ARIMA(1,0,0) with ARCH(2) model of DSES suggest that the actual and fitted data are very close. Therefore, this model fitting is quite reasonable. Figure 4.27 shows the histogram of the estimated residuals from ARIMA(1,0,0) with ARCH(2) model. There is a lack of symmetry in the distribution. When the Jarque-Bera test statistic value is less than or equal to 0.05, the hypothesis of normality is rejected. In Figure 4.27, the Jarque-Bera test statistic value is 7.157 and the p-value is 0.0279. The p-value is less than 0.05, suggesting that the estimated residuals of ARIMA(1,0,0) with the ARCH(2) model of DSES reject the hypothesis of normality. Figure 4.28 shows the correlation of residuals of ARIMA(1,0,0) with ARCH(2) model of DSES. It illustrates that the ACF and PACF of ARIMA(1,0,0) with ARCH(2) model of DSES fall within the 95% confidence interval. All the Prob. values against Q-Stat are greater or equal to 0.05. Therefore, there is no significant autocorrelation at the 5% level.

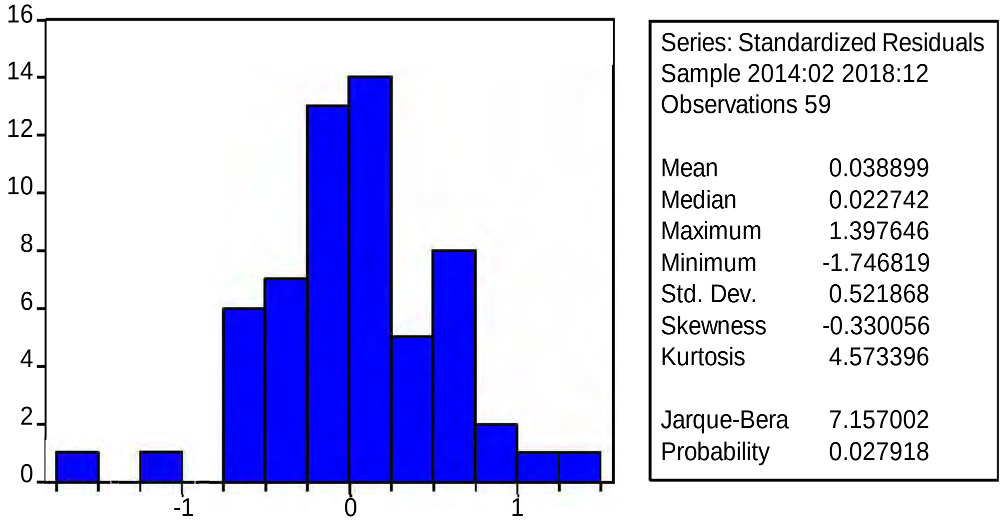


Figure 4.27: Histogram of residuals of ARIMA(1,0,0) with ARCH(2) model of DSES

Sample: 2014:02 2018:12
Included observations: 59
Q-statistic probabilities adjusted for 1 ARMA term(s)

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.240	0.240	3.5819	
		2	-0.059	-0.124	3.8011	0.051
		3	-0.184	-0.149	5.9848	0.050
		4	-0.041	0.040	6.0967	0.107
		5	0.071	0.053	6.4335	0.169
		6	0.055	-0.006	6.6423	0.249
		7	-0.024	-0.034	6.6827	0.351
		8	-0.050	-0.011	6.8575	0.444
		9	0.005	0.029	6.8594	0.552
		10	0.118	0.102	7.8852	0.546
		11	0.178	0.126	10.270	0.417
		12	0.076	0.025	10.715	0.467

Figure 4.28: Correlogram of residuals of ARIMA(1,0,0) with ARCH(2) model of DSES

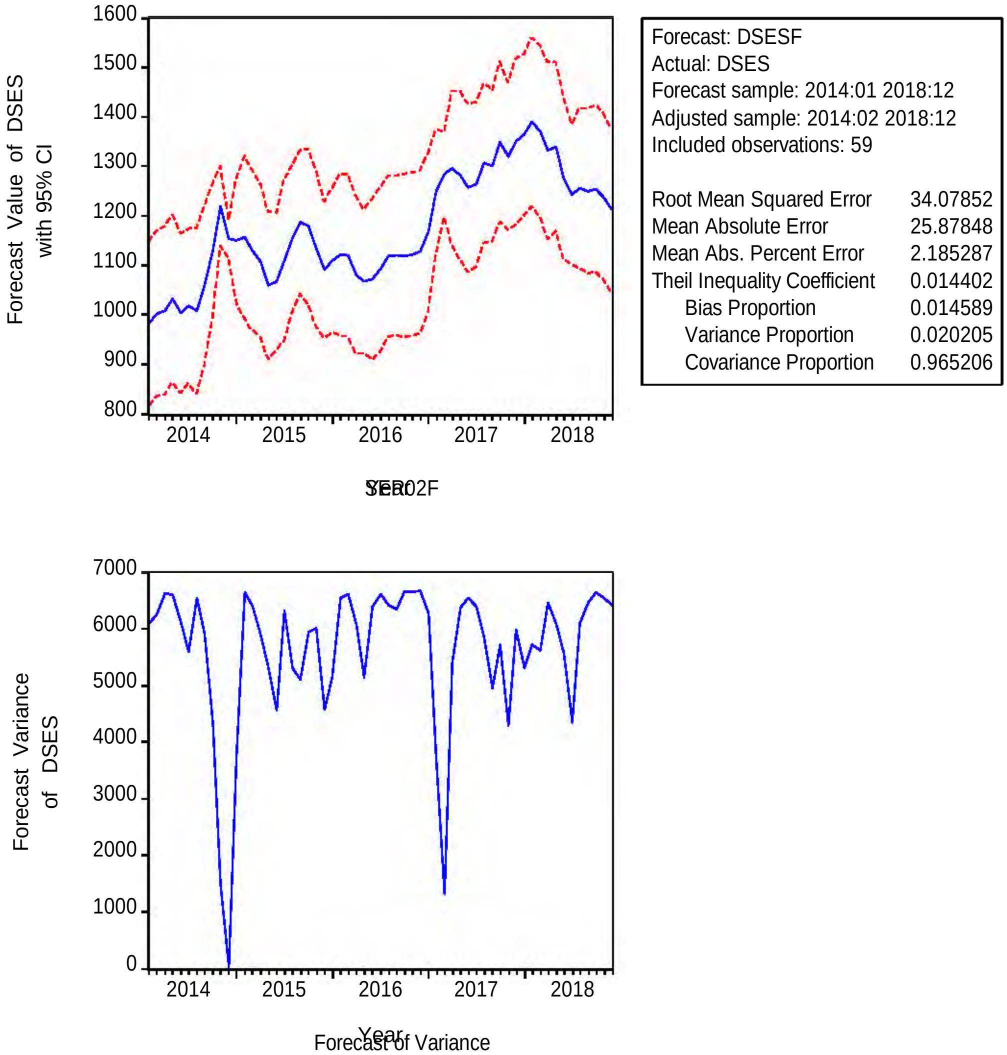


Figure 4.29: Static forecasting performance of ARIMA(1,0,0) with ARCH(2) model of DSES

Figure 4.29 illustrates the static forecasting performance of ARIMA(1,0,0) with ARCH(2) model of DSES. The RMSE of ARIMA(1,0,0) with ARCH(2) model is 34.0785, which is comparatively less than the RMSE of ARIMA and GARCH family models in DSES. The Theil inequality coefficient, biased proportion and variance proportion are approximately close to zero. Therefore, the forecasting performance of ARIMA(1,0,0) with ARCH(2) model of DSES is quite reasonable. The out-of-sample forecast of ARIMA(1,0,0) with ARCH(2) model of DSES from January 2019 to December 2025 is presented in Table 4.31. Static forecasting is conducted for one-step forecast and dynamic forecasting is conducted for the multi-steps forecast.

Table 4.31: Out-of-sample forecast of ARIMA(1,0,0) with ARCH(2) model of DSES

Month	Dynamic Forecast	Static Forecast	Month	Dynamic Forecast	Static Forecast
Jan-19	1174.2280	1215.4208	Jan-22	1178.3707	1165.1768
Feb-19	1174.4400	1212.8492	Feb-22	1178.4760	1163.9002
Mar-19	1174.6377	1210.4513	Mar-22	1178.5812	1162.6236
Apr-19	1174.8221	1208.2155	Apr-22	1178.6865	1161.3470
May-19	1174.9939	1206.1308	May-22	1178.7917	1160.0704
Jun-19	1175.1542	1204.1870	Jun-22	1178.8970	1158.7938
Jul-19	1175.3037	1202.3745	Jul-22	1179.0023	1157.5172
Aug-19	1175.4430	1200.6845	Aug-22	1179.1075	1156.2406
Sep-19	1175.5729	1199.1087	Sep-22	1179.2128	1154.9640
Oct-19	1175.6941	1197.6394	Oct-22	1179.3180	1153.6874
Nov-19	1175.8070	1196.2694	Nov-22	1179.4233	1152.4108
Dec-19	1175.9124	1194.9919	Dec-22	1179.5286	1151.1342
Jan-20	1176.0106	1193.8008	Jan-23	1179.6338	1149.8576
Feb-20	1176.1022	1192.6902	Feb-23	1179.7391	1148.5810
Mar-20	1176.1875	1191.6546	Mar-23	1179.8443	1147.3044
Apr-20	1176.2672	1190.6890	Apr-23	1179.9496	1146.0277
May-20	1176.3414	1189.7887	May-23	1180.0548	1144.7511
Jun-20	1176.4106	1188.9492	Jun-23	1180.1601	1143.4745
Jul-20	1176.4752	1188.1664	Jul-23	1180.2654	1142.1979
Aug-20	1176.5353	1187.4365	Aug-23	1180.3706	1140.9213
Sep-20	1176.5914	1186.7560	Sep-23	1180.4759	1139.6447
Oct-20	1176.6438	1186.1214	Oct-23	1180.5811	1138.3681
Nov-20	1176.6926	1185.5298	Nov-23	1180.6864	1137.0915
Dec-20	1176.7380	1184.9781	Dec-23	1180.7917	1135.8149
Jan-21	1177.1076	1180.4960	Jan-24	1180.8969	1134.5383
Feb-21	1177.2129	1179.2194	Feb-24	1181.0022	1133.2617
Mar-21	1177.3181	1177.9428	Mar-24	1181.1074	1131.9851
Apr-21	1177.4234	1176.6662	Apr-24	1181.2127	1130.7085
May-21	1177.5286	1175.3896	May-24	1181.3180	1129.4319
Jun-21	1177.6339	1174.1130	Jun-24	1181.4232	1128.1553
Jul-21	1177.7392	1172.8364	Jul-24	1181.5285	1126.8787
Aug-21	1177.8444	1171.5598	Aug-24	1181.6337	1125.6021
Sep-21	1177.9497	1170.2832	Sep-24	1181.7390	1124.3255
Oct-21	1178.0549	1169.0066	Oct-24	1181.8442	1123.0489
Nov-21	1178.1602	1167.7300	Nov-24	1181.9495	1121.7723
Dec-21	1178.2655	1166.4534	Dec-24	1182.0548	1120.4957

Month	Dynamic Forecast	Static Forecast	Month	Dynamic Forecast	Static Forecast
Jan-25	1182.1600	1119.2191	Jul-25	1182.7916	1111.5595
Feb-25	1182.2653	1117.9425	Aug-25	1182.8968	1110.2829
Mar-25	1182.3705	1116.6659	Sep-25	1183.0021	1109.0063
Apr-25	1182.4758	1115.3893	Oct-25	1183.1074	1107.7297
May-25	1182.5811	1114.1127	Nov-25	1183.2126	1106.4531
Jun-25	1182.6863	1112.8361	Dec-25	1183.3179	1105.1765

4.5.2.3 ANN Models of DSES

The input structures of various input variables are calculated in this analysis by setting the nodes of the input layer equal to the number of lagged variables from DSES ($x_{t-1}, x_{t-2}, \ldots, x_{t-p}$), where p is a time delay. The input variables for training and testing ANN models of the output variable DSES is $x_t = f\,(x_{t-1}, x_{t-2}, x_{t-3})$. Different ANN models are trained and tested using the software STATISTICA 12. RBF networks and MLP networks are applied. MLP networks perform better than RBF networks for the DSES index. Different training algorithms with MLP networks like BFGS are used. We have applied different types of activation functions like Logistic, Identity, Exponential, Tanh, etc. which are performed as hidden activation and output activation. Summary of ANN models of DSES is exposed in Table 4.32.

Table 4.32: Summary of ANN models of DSES

Net Name	MLP 3-7-1	MLP 3-4-1	MLP 3-8-1	MLP 3-3-1
Training Performance	0.9581	0.9563	0.9558	0.9374
Test Performance	0.9639	0.9599	0.9636	0.9670
Overall Performance	0.9610	0.9581	0.9597	0.9522
Training error	0.0030	0.0031	0.0031	0.0044
Test error	0.0017	0.0019	0.0017	0.0019
Training Algorithm	BFGS 132	BFGS 26	BFGS 17	BFGS 22
Hidden Activation	Tanh	Tanh	Logistic	Identity
Output Activation	Tanh	Exponential	Exponential	Exponential

Note. Output Variable: DSES, Sample (training): 2014:01 2017:12, Sample (test): 2018:01 2018:12

Based on Table 4.32, we observe that MLP 3-7-1's overall performance is better than that of other MLPs. Training and test error of the MLP 3-7-1 net is lower than other MLP nets. So, the out-of-sample forecasts are conducted using MLP 3-7-1 net. The out-of-sample forecast of the ANN model of DSES from January 2019 to December 2025 is presented in Table 4.33.

Table 4.33: Out-of-sample forecast of ANN model of DSES

Month	Forecast of DSES	Month	Forecast of DSES
Jan-19	1177.23	Aug-21	1292.71
Feb-19	1248.98	Sep-21	1292.53
Mar-19	1290.46	Oct-21	1292.36
Apr-19	1307.05	Nov-21	1292.19
May-19	1295.20	Dec-21	1292.02
Jun-19	1266.81	Jan-22	1291.85
Jul-19	1273.76	Feb-22	1291.67
Aug-19	1310.58	Mar-22	1291.50
Sep-19	1314.83	Apr-22	1291.33
Oct-19	1346.32	May-22	1291.16
Nov-19	1335.15	Jun-22	1290.98
Dec-19	1351.46	Jul-22	1290.81
Jan-20	1365.01	Aug-22	1290.64
Feb-20	1378.85	Sep-22	1290.47
Mar-20	1373.16	Oct-22	1290.30
Apr-20	1346.21	Nov-22	1290.12
May-20	1347.63	Dec-22	1289.95
Jun-20	1286.91	Jan-23	1289.78
Jul-20	1249.58	Feb-23	1289.61
Aug-20	1263.80	Mar-23	1289.44
Sep-20	1258.71	Apr-23	1289.26
Oct-20	1263.80	May-23	1289.09
Nov-20	1242.88	Jun-23	1288.92
Dec-20	1211.10	Jul-23	1288.75
Jan-21	1293.91	Aug-23	1288.58
Feb-21	1293.74	Sep-23	1288.40
Mar-21	1293.57	Oct-23	1288.23
Apr-21	1293.39	Nov-23	1288.06
May-21	1293.22	Dec-23	1287.89
Jun-21	1293.05	Jan-24	1287.72
Jul-21	1292.88	Feb-24	1287.54

Month	Forecast of DSES	Month	Forecast of DSES
Mar-24	1287.37	Feb-25	1285.48
Apr-24	1287.20	Mar-25	1285.31
May-24	1287.03	Apr-25	1285.13
Jun-24	1286.85	May-25	1284.96
Jul-24	1286.68	Jun-25	1284.79
Aug-24	1286.51	Jul-25	1284.62
Sep-24	1286.34	Aug-25	1284.45
Oct-24	1286.17	Sep-25	1284.27
Nov-24	1285.99	Oct-25	1284.10
Dec-24	1285.82	Nov-25	1283.93
Jan-25	1285.65	Dec-25	1283.76

4.5.2.4 SVM Models of DSES

The same input structures for the data set are used in the training and testing of SVM models. In this analysis, the input structures of different input variables are calculated by setting the nodes of the input layer equal to the number of the lagged variables from DSES (x_{t-1}, x_{t-2}, ..., x_{t-p}), where p is a time delay The input variables for training and testing SVM models of the output variable DSES is $x_t = f\,(x_{t-1}, x_{t-2}, x_{t-3})$. In the performance of the SVM model, the selection of model and parameter searching play a crucial role. The efficiency of the SVM generalization (estimation accuracy) must depend on a good setting of c, ε, and kernel parameters (Samsudin et al., 2010). This study reflects on the use of the RBF kernel in previous studies for its improved efficiency and advantages in time series forecasting (Ding et al., 2008; Eslamian et al., 2008; Wang et al., 2009). With limited numerical difficulty, the RBF kernel nonlinearly maps samples into a higher dimensional space that can solve nonlinear problems. Different SVM regression models are trained and tested using the software STATISTICA 12. SVM models summary of DSES is shown in Table 4.34.

Table 4.34: SVM models summary of DSES

SVM Type	SVM Parameter		Kernel Type	Kernel Parameter (γ)	No. of support vectors	RMSE		
	C	ε				Train	Test	Overall
Regression	10.0	0.10	RBF	0.333	18	250.549	275.414	262.312
Regression	10.0	0.10	Polynomial (degree=3)	0.333	17	360.357	373.592	363.061
Regression	10.0	0.10	Sigmoid	0.333	41	492.507	482.338	485.585

Note. Output Variable: DSES, Sample (training): 2014:01 2017:12, Sample (test): 2018:01 2018:12

Table 4.34 shows that the RBF kernel type has a lower RMSE than other kernel types. So, regression-based SVM model kernel type RBF is more reasonable for out-of-sample forecasting. The out-of-sample forecast of the SVM model of DSES from January 2019 to December 2025 is presented in Table 4.35.

Table 4.35: Out-of-sample forecast of SVM model of DSES

Month	Forecast of DSES	Month	Forecast of DSES
Jan-19	2447.256	Jun-20	2542.548
Feb-19	2549.164	Jul-20	2512.924
Mar-19	2598.200	Aug-20	2520.816
Apr-19	2599.908	Sep-20	2519.592
May-19	2564.280	Oct-20	2509.552
Jun-19	2537.856	Nov-20	2463.064
Jul-19	2597.848	Dec-20	2430.696
Aug-19	2621.008	Jan-21	2567.978
Sep-19	2685.124	Feb-21	2565.331
Oct-19	2680.024	Mar-21	2562.684
Nov-19	2703.084	Apr-21	2560.037
Dec-19	2735.068	May-21	2557.390
Jan-20	2777.996	Jun-21	2554.743
Feb-20	2765.272	Jul-21	2552.096
Mar-20	2722.624	Aug-21	2549.449
Apr-20	2698.684	Sep-21	2546.802
May-20	2642.964	Oct-21	2544.155

Month	Forecast of DSES	Month	Forecast of DSES
Nov-21	2541.508	Dec-23	2475.334
Dec-21	2538.861	Jan-24	2472.687
Jan-22	2536.214	Feb-24	2470.040
Feb-22	2533.567	Mar-24	2467.393
Mar-22	2530.920	Apr-24	2464.746
Apr-22	2528.273	May-24	2462.099
May-22	2525.626	Jun-24	2459.452
Jun-22	2522.979	Jul-24	2456.805
Jul-22	2520.332	Aug-24	2454.158
Aug-22	2517.685	Sep-24	2451.511
Sep-22	2515.038	Oct-24	2448.864
Oct-22	2512.391	Nov-24	2446.217
Nov-22	2509.744	Dec-24	2443.570
Dec-22	2507.097	Jan-25	2440.923
Jan-23	2504.450	Feb-25	2438.276
Feb-23	2501.803	Mar-25	2435.629
Mar-23	2499.156	Apr-25	2432.982
Apr-23	2496.509	May-25	2430.335
May-23	2493.863	Jun-25	2427.688
Jun-23	2491.216	Jul-25	2425.041
Jul-23	2488.569	Aug-25	2422.394
Aug-23	2485.922	Sep-25	2419.747
Sep-23	2483.275	Oct-25	2417.100
Oct-23	2480.628	Nov-25	2414.454
Nov-23	2477.981	Dec-25	2411.807

4.5.2.5 Comparison of Different Models of DSES

The RMSE for overall samples during January 2014 to December 2018 of the best fitting ARIMA with GARCH family, ANN and SVM model of DSES is presented in Table 4.36.

Table 4.36: RMSE of the estimated models of DSES

Model Name	RMSE
ARIMA with GARCH family	34.0785
ANN	31.6213
SVM	262.312

According to Table 4.36, the RMSE of ANN model is lower than ARIMA with GARCH family model and SVM model of the DSES index. Therefore, the ANN model is the most reliable model for forecasting the DSES index.

4.5.3 Modeling and Forecasting of DSE30

ARIMA, ARIMA with GARCH family, ANN and SVM models of DSE30 index are trained from January 2014 to December 2017 and ARIMA, ARIMA with GARCH family, ANN and SVM models are tested from January 2018 to December 2018.

4.5.3.1 ARIMA Model of DSE30

The The best-fitting ARIMA model is estimated using auto.arima( ) of R Package forecast, Version 8.13 (Rob J. et al., 2020). Even though DSE30 is stationary at level, auto.arima( ) selects the ARIMA(1,1,1) model. Therefore, the ARIMA(1,1,1) model of DSE30 is estimated. Table 4.37 summarizes the ARIMA(1,1,1) model of DSE30 during the training period. There is a significant AR(1) and MA(1) at the 5% level. The RMSE of the test period is greater than the RMSE of the training period. The actual, fitted, and residual plot of the ARIMA(1,1,1) model of DSE30 during the training is shown in Figure 4.30. A shown in Figure 4.30, it is evident that the fitted and actual data points are very different. As a result, out-of-sample forecasts from ARIMA(1,1,1) model are not suitable.

Table 4.37: Summary of ARIMA(1,1,1) model of DSE30 during training period

Variable	Coefficient	Std. Error	t-Statistic	Prob.
c	12.2816	9.1786	1.3381	0.1879
ϕ_1	-0.7854	0.0933	-8.4213	0.0000
θ_1	0.9669	0.0367	26.3555	0.0000
AIC	10.9719			
BIC	11.0912			
R-square	0.13821			
RMSE	57.0930			
RMSE (Test)	65.9893			

Note. Dependent Variable: D(DSE30), Method: LS, Sample (training): 2014:01 2017:12, Sample (testing): 2018:01 2018:12

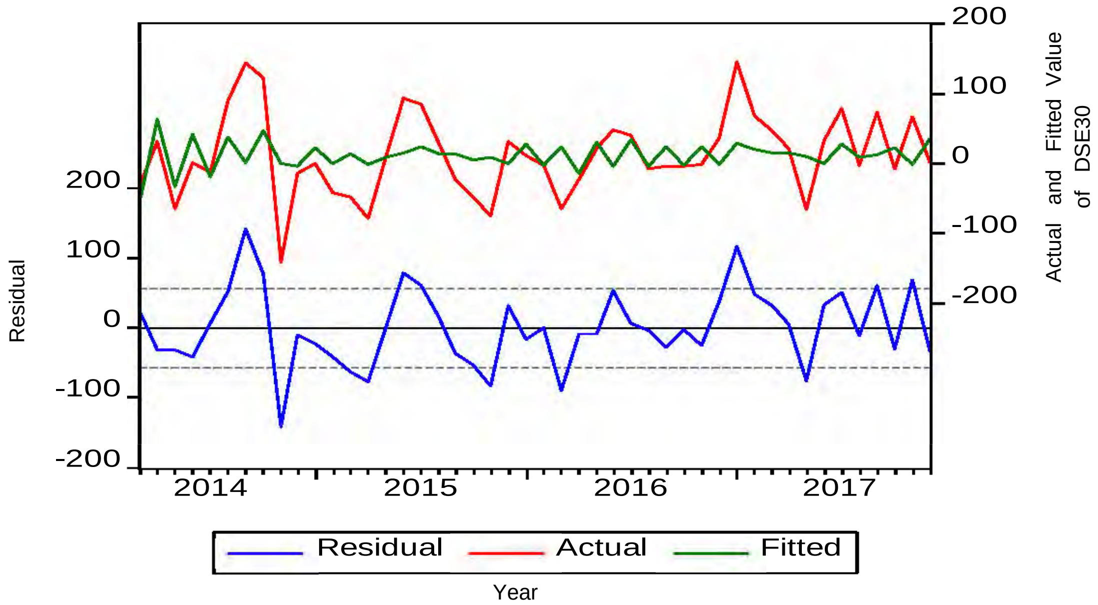


Figure 4.30: Actual, fitted, and residual plot of ARIMA(1,1,1) model of DSE30

4.5.3.2 ARIMA with GARCH Family Model of DSE30

Using ML - ARCH (Marquardt) method, various types of GARCH family models are estimated. To select the best performed GARCH family models, the minimum value of AIC and BIC is considered. The summary of the GARCH family models of DSE30 is presented in Table 4.38. ARCH(2) model has the lowest AIC and BIC value of all models presented in Table 4.38. From the GARCH family of models, the ARCH(2) model is chosen for the DSE30 index. The model ARIMA(1,1,1) does not fit well since DSE30 is stationary at level. Hence, a finite mixture of ARIMA(1,0,1) with ARCH(2) model is selected and then estimated with the DSE30 index. The model summary of ARIMA(1,0,1) with ARCH(2) model of DSE30 is presented in Table 4.39.

Table 4.38: Summary of GARCH family models of DSE30

Models	**Coefficients with constant (Probability)**					**AIC**	**BIC**
ARCH(1)	2642.805*	0.269				11.073	11.152
	(0.001)	(0.298)					
ARCH(2)	**2865.466***	**0.361**	**-0.146**			**10.966**	**11.084**
	(0.000)	**(0.109)**	**(0.371)**				
ARCH(3)	3047.802*	0.258	-0.087	-0.089		10.972	11.129
	(0.001)	(0.076)	0.245	0.581			
GARCH(1,1)	3158.530	0.319	-0.173			11.112	11.230
	(0.131)	(0.237)	(0.737)				
GARCH(2,1)	5051.614*	0.477	-0.371	-0.357*		10.984	11.141
	(0.000)	(0.088)	(0.084)	(0.040)			
GARCH(2,2)	2515.020	0.103	-0.166*	0.433	-0.120	11.096	11.293
	(0.507)	(0.475)	(0.031)	(0.668)	(0.916)		
EGARCH(1,1,1)	8.431	0.049	0.559*	-0.062		11.074	11.232
	(0.065)	(0.917)	(0.036)	(0.913)			
EGARCH(2,1,1)	9.345*	0.123	0.517	0.250	-0.432	11.099	11.296
	(0.005)	(0.789)	(0.065)	(0.534)	(0.386)		

Note. Dependent Variable: D(DSE30), Method: ML - ARCH (Marquardt), Sample (training): 2014:01 2017:12. * indicates significant coefficient at the 5% level.

Table 4.39: Summary of ARIMA(1,0,1) with ARCH(2) model of DSES30

ARIMA(1,0,1) Equation				
	Coefficient	**Std. Error**	**z-Statistic**	**Probability**
c	1928.904*	71.59169	26.94312	0.0000
ϕ_1	0.631340*	0.175346	3.600537	0.0003
θ_1	0.461289*	0.154340	2.988779	0.0028
ARCH(2) Equation				
α_0	20912.15*	10253.01	2.039611	0.0414
α_1	-0.678769	0.774661	-0.876214	0.0809
α_2	-0.661602	0.879537	-0.752217	0.0519

Training sample		**Test sample**	
R-square	0.846	R-square	0.831
AIC	11.962	AIC	10.324
BIC	12.197	BIC	10.526
F-statistic	45.073*	F-statistic	7.922*
Prob(F-statistic)	0.000	Prob(F-statistic)	0.012
RMSE	69.894	RMSE	50.741

Note. Dependent Variable: DSE30, Method: ML - ARCH (Marquardt), Sample (training): 2014:01 2017:12, Sample (test): 2018:01 2018:12. * indicates significant coefficient at the 5% level.

Table 4.39 shows that the R-square values of training and testing periods are 0.840 and 0.866, respectively. A higher R-square value indicates that the models perform well during training and testing. The overall model fitting statistics are significant at the 5% level for training and test samples. Thus, ARIMA(1,0,1) with ARCH(2) is one of the suitable models for out-of-sample forecasting of the DSE30 index. The actual, fitted, and residual plot of ARIMA(1,0,1) with ARCH(2) model of DSE30 index from January 2014 to December 2018 is shown in Figure 4.31.

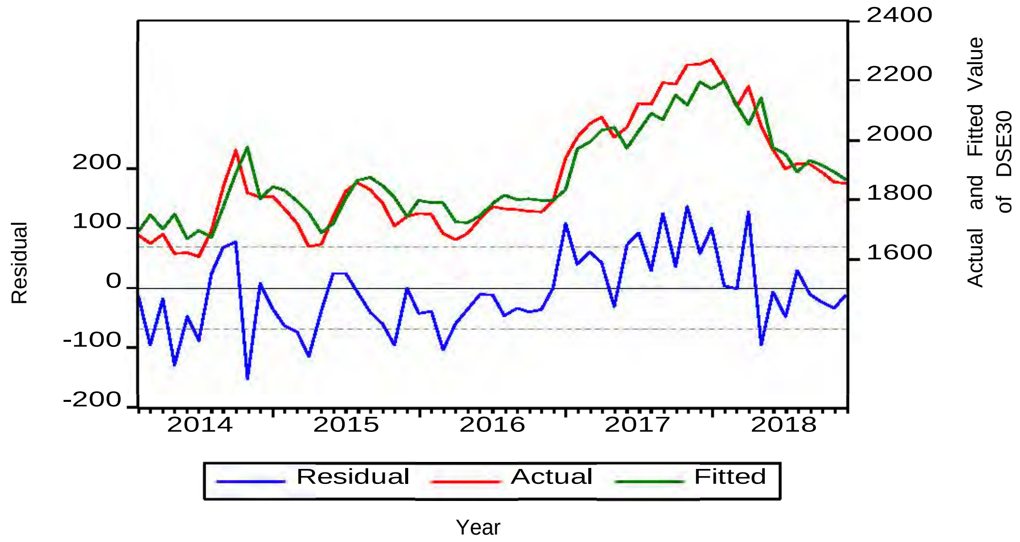


Figure 4.31: Actual, fitted, and residual plot of ARIMA(1,0,1) model of DSE30

According to Figure 4.31, the actual, fitted, and residual plot of ARIMA(1,0,1) with ARCH(2) model of DSES suggest that the actual and fitted data are approximately similar. As a result, this modeling fitting is quite reasonable. Figure 4.32 represents the histogram of estimated residuals from ARIMA(1,0,1) with the ARCH(2) model of DSE30. It suggests that the shape of the distribution is close to symmetrical. The Jarque-Bera test rejects the hypothesis of normality when the p-value is less than or equal to 0.05. From Figure 4.32, the Jarque-Bera test statistic value is 3.5922 and the p-value is 0.1659. Since the p-value is greater than 0.05, it suggests that the estimated residuals from the ARIMA(1,0,1) with ARCH(2) model of DSE30 accepts the hypothesis of normality at the 5% level. Figure 4.33 represents the histogram of estimated standardized residuals from ARIMA(1,0,1) with ARCH(2) model of DSE30. The histogram of standardized residuals of ARIMA(1,0,1) with ARCH(2) model of DSE30 is approximately close to normal. Therefore, the error distribution of the ARIMA(1,0,1) with ARCH(2) model of DSE30 is normal. Figure 4.34 shows the static forecasting performance of ARIMA(1,0,1) with ARCH(2) model of DSE30. The RMSE of ARIMA(1,0,1) with ARCH(2) model of DSE30 is 65.299. The Theil inequality coefficient and biased proportion are approximately close to zero. Therefore, the forecasting performance of ARIMA(1,0,1) with the ARCH(2) model of DSE30 is quite reasonable.

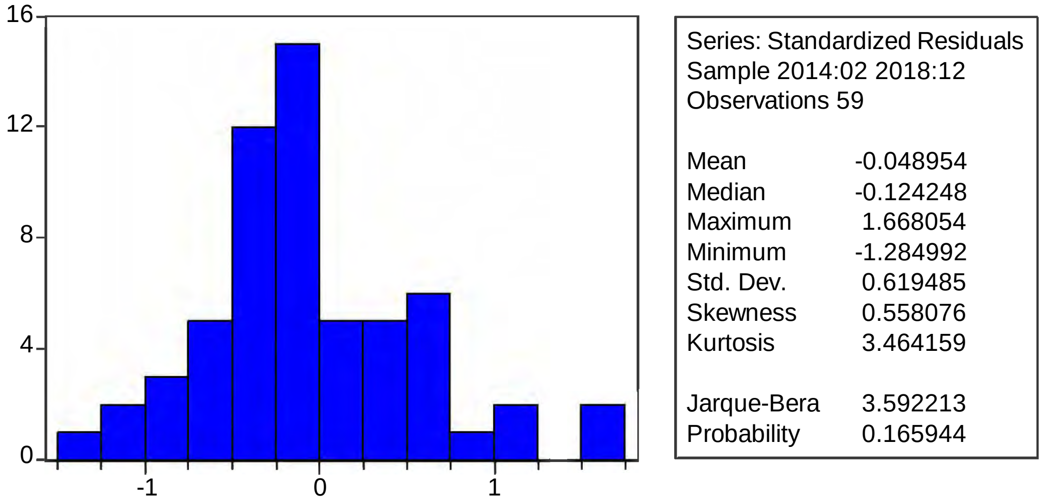


Figure 4.32: Histogram of residuals of ARIMA(1,0,1) with ARCH(2) model of DSE30

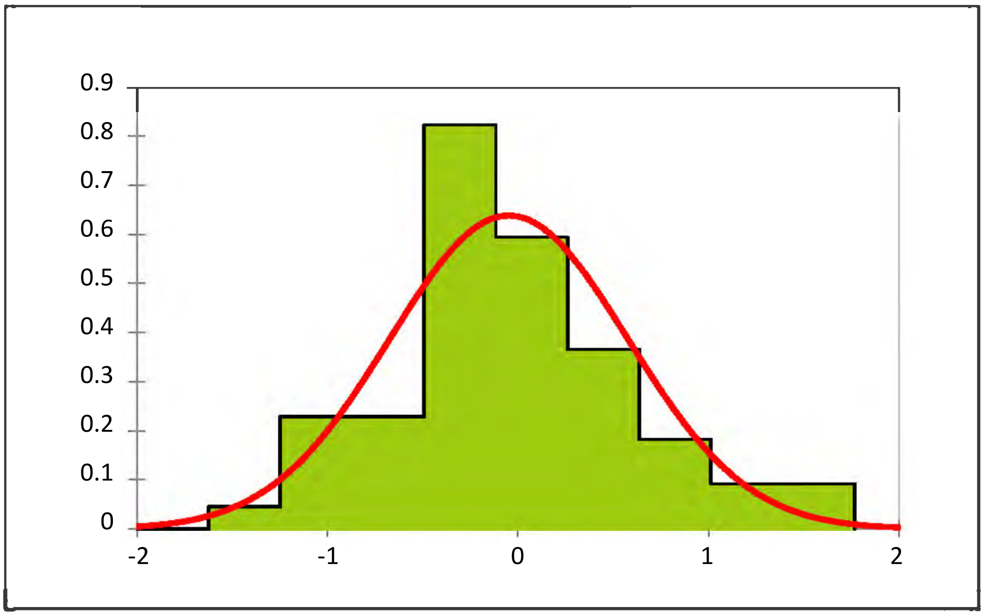


Figure 4.33: Histogram of the standardized residuals of ARIMA(1,0,1) with ARCH(2) model of DSE30

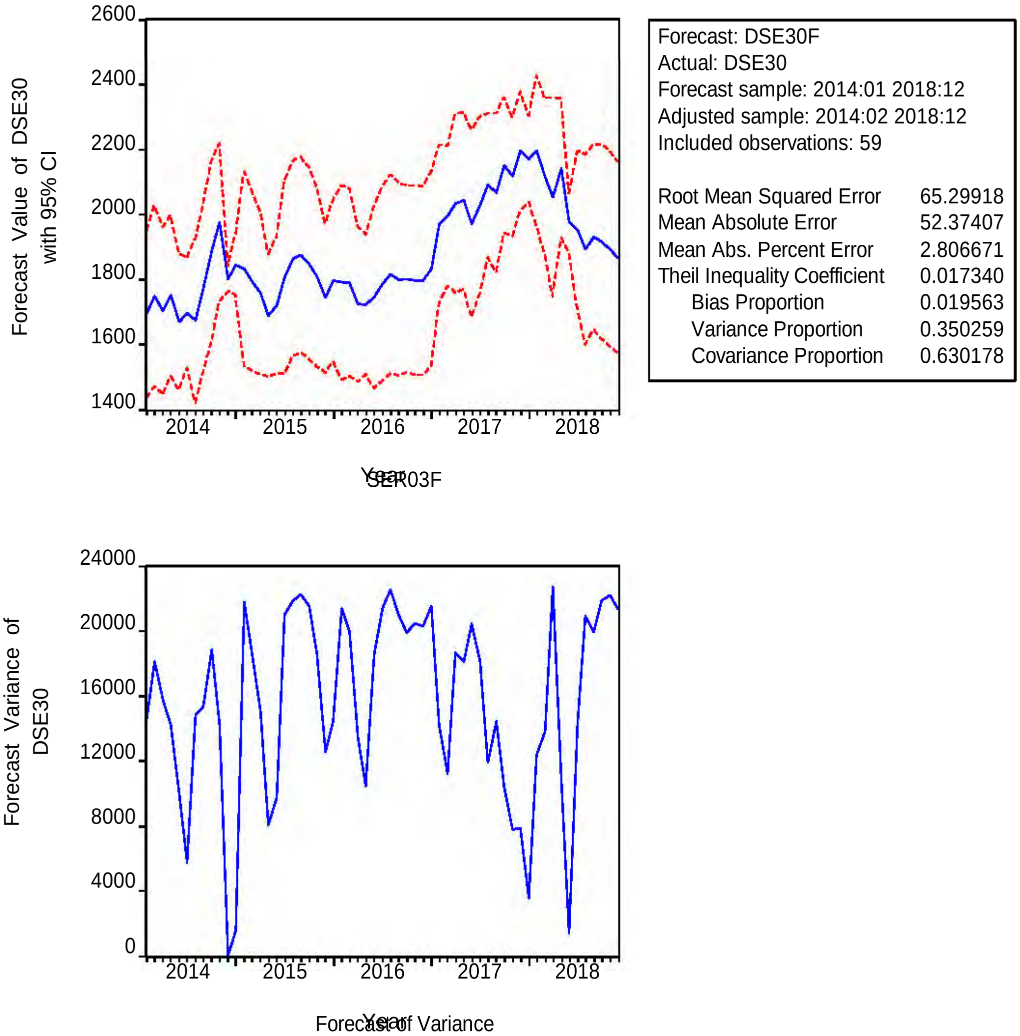


Figure 4.34: Static forecasting performance of ARIMA(1,0,1) with ARCH(2) model of DSE30

The out-of-sample forecast of ARIMA(1,0,1) with ARCH(2) model of DSE30 index from January 2019 to December 2025 is shown in Table 4.40. Static forecasting is conducted for one-step forecast and dynamic forecasting is conducted for the multi-steps forecast.

Table 4.40: Out-of-sample forecast of ARIMA(1,0,1) with ARCH(2) model of DSE30

Month	Dynamic Forecast	Static Forecast	Month	Dynamic Forecast	Static Forecast
Jan-19	1898.4382	1870.4395	Apr-22	1949.6673	1933.7998
Feb-19	1899.7518	1885.6576	May-22	1950.9808	1934.1582
Mar-19	1901.0654	1896.2373	Jun-22	1952.2944	1934.5079
Apr-19	1902.3789	1903.5924	Jul-22	1953.6080	1934.8495
May-19	1903.6925	1908.7057	Aug-22	1954.9215	1935.1833
Jun-19	1905.0060	1912.2605	Sep-22	1956.2351	1935.5096
Jul-19	1906.3196	1914.7318	Oct-22	1957.5486	1935.8288
Aug-19	1907.6332	1916.4499	Nov-22	1958.8622	1936.1413
Sep-19	1908.9467	1917.6443	Dec-22	1960.1758	1936.4471
Oct-19	1910.2603	1918.4746	Jan-23	1961.4893	1936.7467
Nov-19	1911.5739	1919.0519	Feb-23	1962.8029	1937.0404
Dec-19	1912.8874	1919.4532	Mar-23	1964.1165	1937.3282
Jan-20	1914.2010	1919.7323	Apr-23	1965.4300	1937.6105
Feb-20	1915.5146	1919.9262	May-23	1966.7436	1937.8874
Mar-20	1916.8281	1920.0611	Jun-23	1968.0572	1938.1593
Apr-20	1918.1417	1920.1548	Jul-23	1969.3707	1938.4261
May-20	1919.4553	1920.2200	Aug-23	1970.6843	1938.6882
Jun-20	1920.7688	1920.2653	Sep-23	1971.9979	1938.9457
Jul-20	1922.0824	1920.2968	Oct-23	1973.3114	1939.1987
Aug-20	1923.3960	1920.3187	Nov-23	1974.6250	1939.4475
Sep-20	1924.7095	1920.3339	Dec-23	1975.9386	1939.6921
Oct-20	1926.0231	1920.3445	Jan-24	1977.2521	1939.9327
Nov-20	1927.3367	1920.3519	Feb-24	1978.5657	1940.1694
Dec-20	1928.6502	1920.3570	Mar-24	1979.8793	1940.4023
Jan-21	1929.9638	1926.9921	Apr-24	1981.1928	1940.6316
Feb-21	1931.2773	1927.5593	May-24	1982.5064	1940.8574
Mar-21	1932.5909	1928.1052	Jun-24	1983.8199	1941.0797
Apr-21	1933.9045	1928.6314	Jul-24	1985.1335	1941.2988
May-21	1935.2180	1929.1393	Aug-24	1986.4471	1941.5146
Jun-21	1936.5316	1929.6301	Sep-24	1987.7606	1941.7273
Jul-21	1937.8452	1930.1049	Oct-24	1989.0742	1941.9369
Aug-21	1939.1587	1930.5647	Nov-24	1990.3878	1942.1436
Sep-21	1940.4723	1931.0105	Dec-24	1991.7013	1942.3474
Oct-21	1941.7859	1931.4431	Jan-25	1993.0149	1942.5485
Nov-21	1943.0994	1931.8633	Feb-25	1994.3285	1942.7468
Dec-21	1944.4130	1932.2717	Mar-25	1995.6420	1942.9425
Jan-22	1945.7266	1932.6690	Apr-25	1996.9556	1943.1356
Feb-22	1947.0401	1933.0557	May-25	1998.2692	1943.3262
Mar-22	1948.3537	1933.4325	Jun-25	1999.5827	1943.5143

Month	Dynamic Forecast	Static Forecast	Month	Dynamic Forecast	Static Forecast
Jul-25	2000.8963	1943.7001	Oct-25	2004.8370	1944.2438
Aug-25	2002.2099	1943.8836	Nov-25	2006.1506	1944.4206
Sep-25	2003.5234	1944.0648	Dec-25	2007.4641	1944.5953

4.5.3.3 ANN Models of DSE30

The input structures of different input variables are calculated in this analysis by setting the nodes of the input layer equal to the number of lagged variables from the DSE30 index $(x_{t-1}, x_{t-2}, ..., x_{t-p})$, where p is a time delay. The input variables for training and testing ANN models of the output variable DSE30 is $x_t = f\ (x_{t-1}, x_{t-2}, x_{t-3})$. Different ANN models are trained and tested using the software STATISTICA 12. RBF networks and MLP networks are applied. MLP networks performed better than RBF networks for the DSE30 index. Different training algorithms with MLP networks like BFGS are used. We have applied different types of activation functions like Logistic, Identity, Exponential, etc. which are acted as hidden activation and output activation. The summary of the ANN models of DSE30 is presented in Table 4.41.

Table 4.41: Summary of ANN models of DSE30

Net Name	MLP 3-9-1	MLP 3-3-1	MLP 3-6-1	MLP 3-5-1
Training Performance	0.9053	0.9058	0.9051	0.8880
Test Performance	0.9288	0.9379	0.9203	0.9428
Overall Performance	0.9171	0.9218	0.9127	0.9154
Training error	0.0070	0.0069	0.0070	0.0082
Test error	0.0052	0.0048	0.0060	0.0052
Training Algorithm	BFGS 5	BFGS 40	BFGS 23	BFGS 66
Hidden Activation	Identity	Logistic	Exponential	Identity
Output Activation	Logistic	Exponential	Logistic	Exponential

Note. Output Variable: DSE30, Sample (training): 2014:01 2017:12, Sample (test): 2018:01 2018:12

According to Table 4.41, the overall performance of MLP 3-3-1 is superior to other MLP networks. Compared to other MLP nets, the MLP 3-3-1 net has the smallest training and testing error. So, the out-of-sample forecasts are conducted using MLP 3-3-1 net. The out-of-sample forecast of ANN model of the DSE30 index from January 2019 to December 2025 is exposed in Table 4.42.

Table 4.42: Out-of-sample forecast of ANN model of DSE30

Month	Forecast of DSE30	Month	Forecast of DSE30
Jan-19	4782.438	Dec-21	6115.368
Feb-19	4355.819	Jan-22	6166.743
Mar-19	4593.273	Feb-22	6218.117
Apr-19	4841.041	Mar-22	6269.491
May-19	4816.496	Apr-22	6320.865
Jun-19	4564.771	May-22	6372.239
Jul-19	4677.362	Jun-22	6423.614
Aug-19	4636.915	Jul-22	6474.988
Sep-19	4578.915	Aug-22	6526.362
Oct-19	4432.238	Sep-22	6577.736
Nov-19	4467.781	Oct-22	6629.111
Dec-19	4502.619	Nov-22	6680.485
Jan-20	4606.301	Dec-22	6731.859
Feb-20	4616.359	Jan-23	6783.233
Mar-20	4725.494	Feb-23	6834.608
Apr-20	4698.387	Mar-23	6885.982
May-20	4857.245	Apr-23	6937.356
Jun-20	5152.468	May-23	6988.730
Jul-20	5641.304	Jun-23	7040.104
Aug-20	5712.808	Jul-23	7091.479
Sep-20	5545.141	Aug-23	7142.853
Oct-20	5470.598	Sep-23	7194.227
Nov-20	5712.307	Oct-23	7245.601
Dec-20	5805.693	Nov-23	7296.976
Jan-21	5550.252	Dec-23	7348.350
Feb-21	5601.626	Jan-24	7399.724
Mar-21	5653.000	Feb-24	7451.098
Apr-21	5704.374	Mar-24	7502.473
May-21	5755.749	Apr-24	7553.847
Jun-21	5807.123	May-24	7605.221
Jul-21	5858.497	Jun-24	7656.595
Aug-21	5909.871	Jul-24	7707.970
Sep-21	5961.246	Aug-24	7759.344
Oct-21	6012.620	Sep-24	7810.718
Nov-21	6063.994	Oct-24	7862.092

Month	Forecast of DSE30	Month	Forecast of DSE30
Nov-24	7913.466	Jun-25	8273.086
Dec-24	7964.841	Jul-25	8324.460
Jan-25	8016.215	Aug-25	8375.835
Feb-25	8067.589	Sep-25	8427.209
Mar-25	8118.963	Oct-25	8478.583
Apr-25	8170.338	Nov-25	8529.957
May-25	8221.712	Dec-25	8581.331

4.5.3.4 SVM models of DSE30

The same input arrangements for the data set are used in the training and testing of SVM models. The input structures of different input variables are calculated in this analysis by setting the input layer nodes equal to the number of the lagged variables from DSE30 (x_{t-1}, x_{t-2}, ..., x_{t-p}), where p is a time delay. The input variables for training and testing SVM models of the output variable DSE30 is $x_t = f(x_{t-1}, x_{t-2}, x_{t-3})$. In the efficiency of the SVM model, model selection and parameter searching play an important role. Therefore, the efficiency of the SVM generalization (estimation accuracy) must depend on a good setting of hyper-parameters c, ε, and kernel parameters (Samsudin et al., 2010). This study reflects on the use of the RBF kernel in previous studies for its improved efficiency and benefits in time series forecasting (Ding et al., 2008; Eslamian et al., 2008; Wang et al., 2009). With limited computational difficulty, the RBF kernel nonlinearly maps samples into a higher dimensional space that can accommodate nonlinear problems. Different SVM regression models are trained and tested using the software STATISTICA 12. SVM models summary of DSE30 is shown in Table 4.43.

Table 4.43: SVM models summary of DSE30

SVM Type	SVM Parameter		Kernel Type	Kernel Parameter (γ)	No. of support vectors	RMSE		
	C	ε				Train	Test	Overall
Regression	10.0	0.10	Linear	-	18	171.884	183.359	174.924
Regression	10.0	0.10	Polynomial (degree=3)	0.333	29	301.623	242.487	287.498
Regression	10.0	0.10	RBF	0.333	21	171.779	194.503	177.935
Regression	10.0	0.10	Sigmoid	0.333	36	957.278	1066.923	986.803

Note. Output Variable: DSE30, Sample (training): 2014:01 2017:12, Sample (test): 2018:01 2018:12

According to Table 4.43, the kernel type linear has a lower RMSE than other kernel types. So, regression-based SVM model kernel type linear is more reasonable for out-of-sample forecasting. The out-of-sample forecast of the SVM model of the DSE30 index from January 2019 to December 2025 is presented in Table 4.44.

Table 4.44: Out-of-sample forecast of SVM model of DSE30

Month	Forecast of DSE30	Month	Forecast of DSE30
Jan-19	4521.063	Apr-20	6046.723
Feb-19	4643.842	May-20	6020.328
Mar-19	4646.789	Jun-20	6199.341
Apr-19	4764.422	Jul-20	6139.381
May-19	4727.919	Aug-20	6003.506
Jun-19	4890.316	Sep-20	5834.328
Jul-19	5155.827	Oct-20	5634.592
Aug-19	5525.336	Nov-20	5710.220
Sep-19	5606.919	Dec-20	5362.971
Oct-19	5689.234	Jan-21	6220.423
Nov-19	5483.517	Feb-21	6279.549
Dec-19	5417.799	Mar-21	6338.675
Jan-20	5701.408	Apr-21	6397.801
Feb-20	5814.072	May-21	6456.927
Mar-20	6012.489	Jun-21	6516.053

Month	Forecast of DSE30	Month	Forecast of DSE30
Jul-21	6575.179	Oct-23	8171.582
Aug-21	6634.305	Nov-23	8230.708
Sep-21	6693.431	Dec-23	8289.834
Oct-21	6752.557	Jan-24	8348.960
Nov-21	6811.683	Feb-24	8408.087
Dec-21	6870.810	Mar-24	8467.213
Jan-22	6929.936	Apr-24	8526.339
Feb-22	6989.062	May-24	8585.465
Mar-22	7048.188	Jun-24	8644.591
Apr-22	7107.314	Jul-24	8703.717
May-22	7166.440	Aug-24	8762.843
Jun-22	7225.566	Sep-24	8821.969
Jul-22	7284.692	Oct-24	8881.095
Aug-22	7343.818	Nov-24	8940.221
Sep-22	7402.944	Dec-24	8999.347
Oct-22	7462.070	Jan-25	9058.473
Nov-22	7521.196	Feb-25	9117.599
Dec-22	7580.322	Mar-25	9176.725
Jan-23	7639.448	Apr-25	9235.851
Feb-23	7698.574	May-25	9294.977
Mar-23	7757.700	Jun-25	9354.103
Apr-23	7816.826	Jul-25	9413.229
May-23	7875.952	Aug-25	9472.355
Jun-23	7935.078	Sep-25	9531.481
Jul-23	7994.204	Oct-25	9590.607
Aug-23	8053.330	Nov-25	9649.733
Sep-23	8112.456	Dec-25	9708.859

4.5.3.5 Comparison of Different Models of DSE30

The RMSE for overall samples from January 2014 to December 2018 of the best fitting ARIMA with GARCH family, ANN and SVM model of DSE30 is shown in Table 4.45.

Table 4.45: RMSE of the estimated models of DSE30

Model Name	RMSE
ARIMA with GARCH family	65.299
ANN	3063.121
SVM	174.924

Table 4.45 shows that the RMSE of the ARIMA with GARCH family model is lower than ANN and SVM models for the DSE30 index. Therefore, ARIMA with GARCH family is the most reliable model for forecasting the DSE30 index.

4.5.4 Comparative Discussion of Proposed Models

The forecasting performance of DSEX, DSES, and DSE30 indices using ARIMA with GARCH family, ANN and SVM models are compared with the key models reviewed in chapter two. The comparative discussion of the forecasting performance of models is given in the following Table 4.46.

Table 4.46: Comparative results of proposed models

Authors	Results
Tay and Cao, 2001	They applied ANN and SVM models using the financial time series of the Chicago Mercantile Market for forecasting. The results showed that SVM performed better than ANN models based on the criteria of normalized mean square error. The performance of forecasted values was tested up to July 1996 using approximately 25% data as the test sample. Out-of-sample forecasting was absent here.
Kim, 2003	This study was to predict the directions of daily change of the stock price index using the SVM model. The results showed that SVM provides a hopeful alternative to stock market forecasts. The performance of forecasted values was tested up to December 1998 using the last 20% data as the test sample. An out-of-sample forecast was not performed here.

Authors	Results
Thissen et al. 2003	They used ARMA, Elman networks (A form of ANN), and SVM models for a data set of a chemometrics study. Using the same test data set for ARMA, Elman networks, and SVM techniques, it observed that forecasting results were equally well for both the SVM and the ARMA model while the Elman network could not predict the series accurately. The performance of forecasted values was tested with only a few data of different time series. In this study, out-of-sample forecasting was untouched as well.
Kaastra and Boyd, 1995	The ANN and the ARIMA models were used to predict future trading volume time series and found that the forecasting performance of ANN model was better than ARIMA model. Out-of-sample forecasting was not conducted here.
Hans and Kasper, 1998	They forecasted foreign currency exchange rates using ANN models. The results showed that ANN models provide a disposing alternative to foreign currency exchange rates prediction than a linear model. Since ANN can generalize from past experience, they characterized a significant advancement over traditional trading systems. Here, the out-of-sample forecasting was not evaluated.
Bhardwaj et al., 2014	They used time series models which were non-structural-mechanical in nature. The ARIMA and GARCH models were studied and applied for modeling and forecasting of spot prices of Gram at

Authors	Results
Bhardwaj et al., 2014 (Continued)	the Delhi market. It was found that the ARIMA model did not capture the volatility present in the data set whereas the GARCH model successfully captured the volatility. The performance of forecasted values was tested with the last 30 observations as the test sample. Here, out-of-sample forecasting was not shown.
The Present Study	In this study, a finite mixture of ARIMA with GARCH family model performs better than linear model (only ARIMA), non-linear model (only GARCH family), ANN and SVM model based on the criteria of RMSE for forecasting DSEX index. But, the ANN model performs better than the SVM model for forecasting DSEX. These results differ from the study of Tay and Cao, 2001; Kim, 2003 results and partially agree with Thissen et al., 2003 results. This study reveals that a finite mixture of ARIMA(1,0,0) with EGARCH(1,1,2) is the most reliable and reasonable model for forecasting the DSEX index of DSE. But, the finite mixture of ARIMA with GARCH family model (RMSE = 34.0785) and ANN model (RMSE = 31.6213) performs approximately equally likely better than the SVM model (RMSE = 262.312) for forecasting the DSES based on RMSE statistics. The forecasting performance of the SVM model is not reasonable. These results support Hans and Kasper's, 1998 results, Kaastra and Boyd's,

Authors	Results
The present study (Continued)	1995 results, and disagree with Tay and Cao's, 2001 and Kim's, 2003 results. This study exposes that ANN (MLP 3-7-1 net) model is the most reliable model for forecasting the DSES index of DSE. On the other hand, the ARIMA with GARCH family (RMSE = 65.299) and the SVM model (RMSE = 174.924) perform better than the ANN model (RMSE = 3063.121) for forecasting the DSE30 index. These results support Tay and Cao's, 2001 and Kim's, 2003 results. The RMSE of the ARIMA with GARCH family model is lower than the ANN and SVM models of the DSE30 index. Therefore, the ARIMA with GARCH family is the most reliable model for forecasting the DSE30 index. This study reveals that a finite mixture of ARIMA(1,0,1) with ARCH(2) model is the most reliable and reasonable model for forecasting the DSE30 index of DSE. To find the suitable model for forecasting DSEX, DSES, and DSE30 indices, we run the models with approximately 75% observations as the training sample and the last 25% observations as the test sample. The out-of-sample forecasts are conducted by using the forecasting techniques (models) with the smallest RMSE for the test set on the original data set. Out-of-sample forecasts are conducted for

Authors	Results
The present study (Continued)	DSEX, DSES, and DSE30 indices using the proposed models up to December 2025.

As reported in Table 4.46, this study concludes that different types of models describe data well for three different DSE indicators, namely, the DSEX, the DSES, and the DSE30 index. This study also argues that a suitable finite mixture of ARIMA with the GARCH family models may perform better than ANN and SVM models. Most of the reviewed papers argue that ANN and SVM models perform better than ARIMA and the GARCH family models. It may be one reason for the volatility of the DSE indices. In the present study, the out-of-sample forecasted values of DSEX, DSES, and DSE30 indices using the proposed models are conducted up to December 2025 while the authors of previous literature were not provided the out-of-sample forecasted values of the desired indicators. Using the out-of-sample forecasted values of DSEX, DSES and DSE30 indices up to December 2025, the investors and policymakers of DSE can be benefited through investment and policymaking.

Chapter Five

Conclusion and Recommendation

This chapter presents the concluding summary of this dissertation. The methodology is applied by generating the proposed decision support system and performing proper models of the selected indicators of DSE. The key models and forecasting metrics are summarized in this chapter. This chapter concludes this study by reflecting on the results found in the previous chapter. An answer to the research question is given in the context of the modeling and forecasting results. In addition, suggestions and policy recommendations are discussed as well.

5.1 Modeling and Forecasting Selected Indicators of DSE

Before modeling and forecasting selected indicators of DSE, EDA was done using a time series plot. STR is directly and immediately affected by microeconomic indicators such as IMC (US$) and TEC. The time series plot of microeconomic indicators of DSE from 1990 to 2012 showed that there were upward trends of STR, TEC, and IMC. The rising scenario in SRT was found in 1999 and 2009; the number of TEC was gradually increasing from 1990 to 2009 and then gradually up and down after 2009. The time series plot shows that from 2005 to 2012, the macroeconomic indicators of the DSE, as well as the GDP, GNI, and GS, were gradually increasing, while the GI, DIR, and GFI levels were rising and declining, respectively. The time series plot of invested stock market capital in Taka (mn), DGI, stock trade, stock volume, and current market value in Taka (mn) for the period of June 2004 to July 2013 showed that each series rose in 2010, except stock volume and there were severe volatility from 2010 to till the end of the day in stock market capital and DGI, stock trade, stock volume, and current market value data series. The time series plot of the newest indicators of DSE from January 2014 to December 2018 showed that there were upward trends of DSEX, DSES, and DSE30 indices, respectively. However, after July 2017, DSES, DSE30, and DSEX indexes started to fall gradually.

5.2 Cobb-Douglas (CD) Functional Regression Analysis

To investigate the direct and immediate impact on the portfolios of DSE prices, CD functional regression form was used considering the output level STR as a dependent variable and the IMC and TEC of DSE as the independent variables. The intercept and slope coefficients of all explanatory variables were statistically significant at least at the 5% level. Overall there was a negative STR due to the fixed cause of the constant, C = -25.00805 and the relationship of STR with TEC was positive (3.842619) and with IMC was also positive (0.363038) from the period 1990 to 2012. Moreover, F-statistic = 90.02 and Prob. value = 0.000 implied that the regression model significantly fits the data. Finally, the R-square value indicated that about 77.5131 percent variation of STR was explained by the explanatory variables—IMC and TEC of DSE. VIF and TV concluded that there was no presence of multicollinearity between IMC and TEC. The error distribution of the estimated residual was normal. The elasticity of STR in DSE with respect to IMC and TEC was a constant return to scale.

5.3 Multiple Linear Regression Analysis

To investigate the indirect and long-run impact on the portfolios of DSE prices, the multiple log-linear regression model was used considering the output level DGI as the dependent variable and the macroeconomic indicators like GDP, GNI, GS, GI, DIR, and GFI, respectively as the independent variables. The R-square value from the model was 0.9998 and overall model fitting was statistically significant at 5% level. The error distribution of the model was normal. But, there were multicollinearity problems. To get rid of this problem, the multiple linear regression model was re-estimated by dropping GNI due to very severe multicollinearity, and standardized GDP, standardized GS, and standardized GFI were used as the explanatory variables due to severe/moderate multicollinearity.

We concluded that the overall model fitting was significant at 10% level and a higher value of R-square (0.973) was found.

5.4 VAR Modeling and Forecasting

VAR Model was trained and tested with the data series like stock trade, invested capital, stock volume, current market value, and DGI. Before building a suitable VAR model, summary statistics were evaluated and outliers were checked by Box and Whisker plots. From Box and Whisker plots, outliers were found in stock trade, volume, and value series. To check the stationary of the series, unit root tests were applied. The unit root tests exposed that market capital, DGI, value, and trade series were non-stationary by not rejecting the null hypothesis of unit-root at 5% levels of significance and critical values, but they were all stationary after first differencing except volume data of DSE which was normally stationary. AIC and BIC values were used to select the lag length of the VAR model. The minimum value of AIC and BIC was found at the lag length of order two than that of any other lag lengths of orders. So, the VAR(2) model was estimated. The VAR(2) model satisfied the stability condition. The distribution of estimated residuals from the VAR(2) model was a lack of multivariate normal distribution. VAR residual heteroscedasticity tests revealed the rejection of the null hypothesis of no ARCH effects. Granger causality test results revealed that there were significant bivariate causal relationships among the variables at 5% level except for value, trade, and volume series to each other. Data from June 2004 to June 2012 were used for training samples and from July 2012 to July 2013 were used for testing samples and compared the results of the VAR(2) model with the univariate auto ARIMA (1,1,1) models. RMSE statistics for overall samples of the VAR(2) model is minimal from ARIMA (1,1,1) models for market capital, DGI, and volume data series of DSE. Therefore, the forecasting performance of the VAR(2) model was more reasonable than ARIMA (1,1,1) models.

5.5 Univariate Modeling and Forecasting

ARIMA, ARIMA with GARCH family, ANN, and SVM model were estimated and analyzed with DSEX, DSES, and DSE30 indices time series. KPSS test was used to check the stationary condition of DSEX, DSES, and DSE30 indices. KPSS tests of DSEX, DSES, and DSE30 indices revealed that DSEX, DSES, and DSE30 indices were stationary at the 5% level.

5.5.1 Modeling and Forecasting of DSEX

Firstly, an auto ARIMA(1,1,0) model of DSEX was estimated. But, the model fitting was not good. Secondly, the best fitting GARCH family model was selected using AIC and BIC values. EGARCH(1,1,2) model was selected for the lowest value of AIC and BIC. Since DSEX was stationary at level and ARIMA(1,1,0) model was not well fitted so, the finite mixture of the ARIMA(1,0,0) with EGARCH(1,1,2) model was selected and then estimated with the DSEX index. The R-square value of the training period was 0.919 and the R-square value of the testing period was 0.811, respectively. The error distribution of the ARIMA(1,0,0) with EGARCH(1,1,2) model was normal. The RMSE of ARIMA(1,0,0) with EGARCH(1,1,2) model was 162.50, which was comparatively lower than RMSE of other ARIMA and GARCH family models of DSEX. The Theil inequality coefficient, biased proportion and variance proportion were approximately close to zero. Therefore, the forecasting performance of ARIMA(1,0,0) with EGARCH(1,1,2) model of DSEX is quiet reasonable.

Different ANN models were trained and tested using the DSEX index. RBF networks and MLP networks were applied. Different training algorithms like BFGS and RBFT were used. MLP nets performed better than RBF nets. Though the test error of the RBF net was minimum than the MLP net, training error of MLP net was minimum than the RBF net. So, the out-of-sample forecasts were conducted using the best-performed MLP 3-10-1 net.

Various SVM models were estimated with the DSEX index. RBF was found to have minimum RMSE compared with other kernel types. So, regression-based SVM model kernel type RBF was a more reasonable model for out-of-sample forecasting.

The RMSE for overall samples during January 2014 to December 2018 of the best fitting ARIMA with GARCH family, ANN and SVM models of DSEX concluded that the RMSE of the ARIMA with GARCH family model was lower than ANN and SVM models of DSEX. Therefore, ARIMA with GARCH family model is the most reliable model for forecasting the DSEX index. The out-of-sample forecasted values of DSEX were conducted up to December 2025 using the proposed model ARIMA(1,0,0) with EGARCH(1,1,2) shown in Table 4.22.

5.5.2 Modeling and Forecasting of DSES

Firstly, an auto ARIMA(1,1,0) model of DSES was estimated. The RMSE of the test period was slightly lower than the RMSE of the training period. The actual, fitted, and residual plot suggested that the model was not well fitted. So, the out-of-sample forecast from ARIMA(1,1,0) model was not suitable. To select the best performed GARCH family models, the minimum value of AIC and BIC was considered. The ARCH(2) model was selected based on the value of AIC and BIC. In this case, DSES was stationary and ARIMA(1,1,0) was not well fitted. So, ARIMA(1,0,0) with ARCH(2) model was selected and then estimated with the DSES index. R-square values of training and testing periods were 0.893 and 0.831, respectively. The higher value of R-square ensured that the models performed well in the training and test period. The overall model fitting statistics were significant at the 5% level for training and test samples. Jarque-Bera test concluded that the estimated residuals from the ARIMA(1,0,0) with ARCH(2) model of DSES was a lack of normality. The RMSE of ARIMA(1,0,0) with ARCH(2) model was 34.0785 which was comparatively lower than RMSE of other ARIMA and GARCH family models of DSES. The Theil inequality coefficient, biased proportion, and variance proportion were approximately close to zero. Therefore, the forecasting performance of ARIMA(1,0,0) with ARCH(2) model of DSES is quite reasonable.

Different ANN models were trained and tested using the DSES index. MLP networks performed better than RBF networks. Different training algorithms with MLP networks like BFGS were used. ANN models concluded that the performance of the MLP 3-7-1 net was better than that of other MLP nets. So, the out-of-sample forecasts were conducted using MLP 3-7-1 net.

Based on the SVM models, the kernel type RBF had a lower RMSE than other kernel types. So, regression-based SVM model kernel type RBF is more reasonable for out-of-sample forecasting.

The RMSE for overall samples during January 2014 to December 2018 of the best fitting ARIMA with GARCH family, ANN and SVM model of DSES revealed that RMSE of ANN model was lower than ARIMA with GARCH family model and SVM models of DSES. Therefore, the ANN model is the most reliable model for forecasting DSES. Table 4.33 shows out-of-sample forecasted values of DSES up to December 2025 based on the proposed model ANN (MLP 3-7-1 net).

5.5.3 Modeling and Forecasting of DSE30

Firstly, an auto ARIMA(1,1,1) model of DSE30 was estimated. The coefficient of AR(1) and MA(1) were significant at the 5% level. The RMSE of the test period was greater than the RMSE of the training period. The actual, fitted, and residual plot concluded that the model fitting was not good. To select the best performed GARCH family models, the minimum value of AIC and BIC was considered. The ARCH(2) model was selected based on the value of AIC and BIC. The DSE30 index was stationary at level and the ARIMA(1,1,0) model was not well fitted. So, the finite mixture of ARIMA(1,0,1) with ARCH(2) model was finally selected and then estimated with the DSE30 index. R-square values of the training and testing periods were 0.846 and 0.831, respectively. The overall model fitting statistics were significant at the 5% level for training and test samples. Therefore, ARIMA(1,0,1) with ARCH(2) model is one of the suitable models for out-of-sample forecasting of the DSE30 index. The Jarque-Bera test revealed that the estimated residuals from the ARIMA(1,0,1) with ARCH(2) model of DSE30 were normal. The RMSE of ARIMA(1,0,1) with ARCH(2) model of DSE30 was 65.299. The Theil inequality coefficient and biased proportion were approximately close to zero. Therefore, the forecasting performance of the ARIMA(1,0,1) with ARCH(2) model of DSE30 is quite reasonable.

Different ANN models were trained and tested using the DSE30 index. RBF networks and MLP networks were applied. MLP networks performed better than RBF networks for the DSE30 index. Different training algorithms with MLP networks like BFGS were used. The ANN models revealed that the overall performance of the MLP 3-3-1 net performed better than other MLP nets. Training and test errors of the MLP 3-3-1 net were lower than other MLP nets. So, the out-of-sample forecasts were conducted using MLP 3-3-1 net.

Different SVM regression models were trained and tested using the DSE30 index. There was a minimum RMSE for the linear kernel type compared to the other kernel types based on SVM models. So, a regression-based SVM model kernel type linear model is a more reasonable model for out-of-sample forecasting.

The RMSE for overall samples during January 2014 to December 2018 of the best fitting ARIMA with GARCH family, ANN and SVM model of DSE30 index revealed that RMSE of ARIMA with GARCH family model was minimum than ANN and SVM models of DSE30. Therefore, ARIMA with GARCH family is the most reliable model for forecasting the DSE30 index. The out-of-sample forecasted values of DSE30 were conducted up to December 2025 using the proposed model ARIMA(1,0,1) with ARCH(2) shown in Table 4.40.

5.6 Suggestions and Policy Recommendations

The growth of an economy depends upon a well-functioning stock market. The stock market plays a vital role in acting as an intermediary between shareholders and companies pursuing extra financing for business extension. The major role of a stock market mostly leads to financial growth by aggregating the funds to the finance industry and other enterprises. The stock market delivers a marketplace along with facilities for bringing together the buyers and sellers of shares, promoting just and equitable principles of trade, and protecting the interest of the shareholders. The capitalization of Bangladesh's stock market contributed to 9.2% of nominal GDP in June 2020, compared to 13.5% in 2019. The contribution to GDP touched a record of 28.5% in June 2010 and 4.2% in June 2006 (CEIC Data, 2020). DSE is the major stock market in Bangladesh. The study of DSE indices is a very blazing issue in Bangladesh and it is essential for policy implications. After the market crash in 1996, DSE was performing healthy with its rising DGI. The DGI eventually crashed at its highest point in May 2010. Then shareholders lost their confidence in the stock market. Thus, the optimistic stock market moved to bearish in November 2010, losing 1800 points from December 2010 to January 2011. Millions of shareholders became bankrupt due to this stock market crash. The crash is supposed to provide benefits to the big players in the stock market by artificially manipulating share prices.

This dissertation established that ARIMA(1,0,0) with EGARCH(1,1,2) model is the most reliable model for forecasting DSEX index, ANN (MLP 3-7-1 net) model is the most reliable model for forecasting DSES index, and ARIMA(1,0,1) with ARCH(2) model is the most reliable model for forecasting DSE30 index. The forecasting of DSEX, DSES, and DSE30 indices with the proposed model was conducted from January 2019 to December 2025. These forecasting results of our study may help BSEC, individual and institutional investors, industry owners, stakeholders, and above all the Government of Bangladesh to take appropriate actions for building an efficient and sustainable stock market in Bangladesh.

The stock index measures the changes in share prices that are generally associated with market conditions. The shareholders consider it as a benchmark to detect stock market conditions with earnings or dividend per share. The market condition of every company somehow depends on the financial condition of the country. The forecasting of DSE indicators may help to determine the stability of the index. Researchers and investors will find this dissertation useful for predicting future share values and making investment decisions by utilizing the modeling and forecasting concept used in this dissertation.

5.7 Further Scopes of Research

In this study, yearly and monthly time series data of DSE are used. We have applied ARIMA, GARCH, ARCH, EGARCH models considering Normal (Gaussian) error distribution. In the future, the researchers can apply Vector Error Correction (VEC), Bayesian VAR, PARCH, TARCH, etc. models and also considering the non-normal error distributions like Student’s t and Generalized Error Distribution (GED) using daily time series indicators of DSE.

Bibliography

Adjasi, C., Harvey, K.S. and Agyapong, D. (2008). Effect of Exchange Rate Volatility on the Ghana Stock Exchange. *African Journal of Accounting, Economics, Finance and Banking Research*, 3(3), pp. 28-47.

Afolabi, M. and Olude, O. (2007). "Predicting Stock Prices Using a Hybrid Kohonen Self Organizing Map (SOM)''. *IEEE 40th Annual Hawaii International Conference: System Sciences*.

Afzal, N. and Hossain, S.S. (2011). An Empirical Analysis of the Relationship between Macroeconomic Variables and Stock Prices in Bangladesh. *Bangladesh Development Studies*, Vol. XXXIV, December 2011, No. 4

Ahmed, E.M.A. and Suliman, Z.S. (2011). Modeling Stock Market Volatility using GARCH models, Evidence from Sudan. *International Journal of Business and Social Science*, 2(23).

Ahmed, M.F. (2000). Emerging Stock Market and the Economy. *South East Asian Studies Series*, Nagasaki University, Japan.

Ahmed, R.A. and Shabri, A.B. (2013). Fitting GARCH models to Crude Oil Spot Price Data. *Life Science Journal*, 10(4).

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions*.

Alam, M.M., Alam, K.A. and Uddin, M.G.S. (2007), Market Depth and Risk Return Analysis of Dhaka Stock Exchange: An Empirical Test of Market Efficiency. *ASA University Review*, Vol. 1(1), pp. 93-101, July-December.

Ali, M.B. (2011). Co integrating Relation between Macroeconomic Variables and Stock Return: Evidence from Dhaka Stock Exchange (DSE). *International Journal of Business and commerce*, 1(2), pp. 25-38.

Ali, S.O. and Mhmoud, A.S. (2013). Estimating Stock Returns Volatility of Khartoum Stock Exchange through GARCH Models. *Journal of American Science*, 9(11).

Ayodele, A. A., Aderemi, O.A. and Charles, K.A. (2014). "Stock Price Prediction Using the ARIMA Model". A paper presented at the 16th International Conference on Computer Modelling and Simulation, University of KwaZulu-Natal Durban, South Africa.

Aziz, I.S. and Uddin, N. (2014). Volatility Estimation in the Dhaka Stock Exchange (DSE) returns by GARCH models. *Asian Business Review*, 4(1).

Bala, D.A. and Asemota, J.O. (2013). Exchange Rate Volatility in Nigeria: Application of GARCH models with Exogeneous Break. *CBN Journal of Applied Statistics*, 4(1).

Banerjee, K.P. and Adhikary, B.K. (2009). Dynamic Effects of Interest Rate and Exchange Rate changes on Stock Market Return in Bangladesh. *Journal of Social and Management Sciences*, Vol. 38, No. 1, pp. 27-40.

Bangladesh Securities and Exchange Commission (BSEC, 2014). *Annual Report 2012-2013*, P 15.

Bangladesh Securities and Exchange Commission (BSEC, 2018). Retrieved from https://sec.gov.bd/

Basel, M., Awartani, A. and Corradi, V. (2005). Predicting the volatility of the S&P-500 Stock index via GARCH models: the Role of Asymmetries. *International Journal of Forecasting*, 21, pp. 167-183.

Bhardwaj, S.P., Ranjit, K.P., Singh, D.R. and Singh, K.N. (2014). An empirical investigation of ARIMA and GARCH models in Agricultural Price Forecasting. *Economic Affairs*, 59(3), pp. 415-428.

Bisgaard, S. and Kulahci, M. (2011). Time Series Analysis and Forecasting by Example. Wiley, A John Wiley & Sons, INC., Publication.

Bollerslev, T. (1986). Generalized Autoregressive Conditional Heteroscedasticity. *Journal of Econometrics*, 31(3), pp. 307-327.

Box, G.E.P. (1976). Science and statistics. *Journal of the American Statistical Association*, 71, pp. 791–799, doi:10.1080/01621459.1976.10480949

Box, G.E.P., and Tiao, G.C. (1975). Intervention analysis with applications to economic and environmental problems. *Journal of the American Statistical Association,* 70 (349), pp.70- 79.

Branes, P. (1986). Thin trading and stock market efficiency: A case of the Kuala Lumpur Stock Exchange. *Journal of Business Finance & Accounting*, 13(4), pp. 609-617.

Burges, C.J.C. (1998). A Tutorial on Support Vector Machines for Pattern Recognition. *Data Mining and Knowledge Discovery*, 2(2), pp. 121-167.

Campbell, J., Lo, A. and MacKinlay, C. (1997). The econometrics of financial markets. Princeton: Princeton University Press.

CEIC Data (2020). Market Capitalization: % of GDP by Country, Bangladesh.

Chan, K.C., Gup, B.E., and Pan, M.S. (1992). An Empirical Analysis of Stock Prices in Major Asian Markets and the United States. *The Financial Review*, Vol. 27(2), pp.89- 307.

Chen, S. et al. (2014). The time series forecasting: from the aspect of network. *Automatic Control*, AC-19, pp. 716 – 723.

Cheung, Y.L., Wong, K. A. and Ho, Y.K. (1993). The pricing of risky assets in two emerging Asian markets-Korea and Taiwan. *Applied Financial Economics*, Vol. 3(4), pp. 315-324.

Chowdhury, A.R. (1994). Statistical Properties of Daily Returns from Dhaka Stock Exchange. *The Bangladesh Development Studies*, 22(4), pp. 61-76.

Claessens, S., Dasgupta, S. and Glen, J. (1995). Return Behaviour in Emerging Stock Markets. *The World Bank Economic Review*, 9(1), pp.131–151.

Cook, R. and Weisberg, S. (2009). An Introduction to Regression Graphics. *Journal of the American Statistical Association*, 90. doi: 10.2307/2291355.

Corhay, A., and Rad, T. (1994). Statistical Properties of Daily Return: Evidence from European Stock Market. *Journal of Business Finance and Accounting*, 21, pp. 271-282.

Culbertson, K. (1996). Quantitative financial economics: Stocks, bonds and foreign exchange. New York: John Wiley & Sons.

Dhaka Stock Exchange Limited (DSE Ltd. Web portal, 2019). Retrieved from https://www.dsebd.org/index.php

Dhaka Stock Exchange Limited (DSE, 2018). Annual Report 2017-2018, pp. 7-9

Dickey, D.A. and Fuller, W.A. (1979). Distribution of the estimates for autoregressive time series with a unit root. *Journal of American Statistical Association,* 74, pp. 427 – 431.

Dickinson, J.P. and Muragu, K. (1994). Market Efficiency in Developing Countries: A case study of the Nairobi Stock Exchange. *Journal of Business Finance & Accounting*, 21(1), pp. 133-150.

Ding, Y., Song, X., and Zen, Y. (2008). Forecasting financial condition of Chinese listed companies based on support vector machine. *Expert Systems with Applications*, 34, pp. 3081-3089. doi: 10.1016/j.eswa.2007.06.037.

Dolado, J.J., Jenkinson, T., Sosvilla-Rivero, S. (1990). Cointegration and Unit Roots. *Journal of Economic Surveys* 4(3), pp. 249–273.

Doruk, Ö. (2016). VAR with first difference and level stationary variables? Retrieved from: https://www.researchgate.net/post/VAR_with_first_difference_and_level_stationary_variables/57bea65a93553b0b4b423e43/citation/download.

Elliott, G., Rothenberg, T.J. and Stock, J.H. (1996). Efficient tests for an autoregressive unit root. *Econometrica*, 64, pp. 813-836.

Engel, R.F. (1982). Autoregressive Conditional Heteroscedasticity with estimates of the variance for U.K. inflation. *Econometrica*, 50, pp. 987-1008.

Eslamian, S., Gohari, A., Biabanaki, M. and Malekian, R. (2008). Estimation of Monthly Pan Evaporation Using Artificial Neural Networks and Support Vector Machines. *Journal of Applied Sciences*, 8, pp. 3497-3502.

Fama, E. (1991). Efficient Capital Markets: II. *Journal of Finance*, 46, pp. 1575-1617. Retrieved from http://dx.doi.org/10.1111/j.1540-6261.1991.tb04636.x

Yule, G.U. (1927). On a Method of Investigating Periodicities in Disturbed Series. *Philosophical Transactions of the Royal Society of London*, 226, pp. 267-298. http://dx.doi.org/10.1098/rsta.1927.000

Hamilton, J.D. (1994a). Time series analysis. Princeton: Princeton University Press.

Hamilton, J.D. (1994b). State space models. In R. F. Engle and D. L. McFadden (Eds.), Handbook of econometrics. Amsterdam: Elsevier.

Hans, F.P. and Kasper, V.G. (1998). Forecasting Exchange Rates Using Neural Networks for Technical Trading Rules. *Studies in Nonlinear Dynamics & Econometrics*, De Gruyter, 2(4), pp. 1-8.

Hansen, P.R. and Lung, A. (2001). A Comparison of Volatility Models: Does Anything Beat a GARCH (1,1)? Centre for Analytical Finance, University of Aarhus, Working Paper Series, 84. http://dx.doi.org/10.2139/ssrn.264571

Haque, M.S., Eunus, R. and Ahmed, M. (2001). Risk Return & Market Efficiency in Capital Market under Distress: Theory and Evidence from DSE. *Chittagong Stock Exchange Publication*, Quarter-1. Available at http://www.csebd.com/cse/Publications/portfolio_Q1_2001/risk_%20return.html, (Accessed on March 22, 2018).

Harvey, C.R. (1994). Conditional Asset allocation in Emerging Markets. Working Paper No.4623, Cambridge, MA.

Hasan, M.A. (2015). Testing weak-form market efficiency of Dhaka stock exchange. *Global Disclosure of Economics and Business*, 4(2), pp.79-90.

Hassan, M.K. and Chowdhury, S.S.H. (2008). Efficiency of Bangladesh stock market: Evidence from monthly index and individual firm data. *Applied Financial Economics*, 18(9), pp. 749-758.

Hassan, M.K., Islam, M.A and Basher, S.A. (1999). "Market Efficiency, Time-Varying Volatility and Equity Returns in Bangladesh Stock Market", Working Papers 2002-06, York University, Department of Economics, revised Jun 2002.

Hossain, M.F. and Nasrin, S. (2012). Factors affecting selection of equity shares: The case of retail investors in Bangladesh. *European Journal of Business and Management*, 4(20), pp. 110-124.

Hossain, M.S. and Kamal, K.M.M. (2010). Does stock market development cause economic growth? A time series analysis for Bangladesh economy. In international conference on applied economics-ICOAE.

Hossain, A., Kamruzzaman, M. and Ali, M.A. (2015). Vector Autoregressive (VAR) Modeling and Projection of DSE. *Chinese Business Review*, Vol. 14, No. 6. doi: 10.17265/1537-1506/2015.06.001

Hossain, A., Kamruzzaman, M. and Ali, M.A. (2016). Econometric Investigation of DSE's Portfolios through Selective Micro and Macro Economic Indicators. *Journal of Statistical Science and Application*, March 2016, Vol. 4, No. 05-06. doi: 10.17265/2328-224X/2016.0506.003

Hossain, A., Kamruzzaman, M. and Ali, M.A. (2015). ARIMA with GARCH Family Modeling and Projection on Share Volume of DSE. *Economics World*, Vol. 3, No. 7-8. doi: 10.17265/2328-7144/2015.0708.003

Hou, Y. and Li, S. (2015). Volatility behaviour of stock index futures in China: a bivariate GARCH approach. *Studies in Economics and Finance*, 32(1), pp. 128-154.

Huang, W., Nakamori, Y. and Wang, S. (2005). Forecasting stock market movement direction with support vector machine. *Computers & Operations Research*, 32(10), pp. 2513-2522.

Investopedia (2018). What is GDP? Retrieved from https://www.investopedia.com/terms/g/gdp.asp

Investopedia (2018). What is GNI? Retrieved from https://www.investopedia.com/terms/g/gross-national-income-gni.asp

Islam, A. and Khaled, M. (2005). Tests of Weak-Form Efficiency of The Dhaka Stock Exchange. *Journal of Business Finance & Accounting*, 32(7-8), pp. 1613-1624.

Islam, M.N. (2011). Problems and prospects of stock market in Bangladesh. *Economic Research*, 12(81).

Jarque, C.M. and Bera, A.K. (1987). A test for normality of observations and regression residuals. *International Statistical Review*, 55 (2), pp. 163–172.

Jenkins, G.M. and Watts, D.G. (1968). Spectral Analysis and Its Applications Holden-Day. San Francisco, pp. 243-238.

Joshi, P. (2010). Modeling Volatility in Emerging stock markets of India and China. *Journal of Quantitative Economics*, 8(1).

Kaastra, I. and Boyd, M.S. (1995). Forecasting futures trading volume using neural networks. *The Journal of Futures Markets*, 15(8), pp. 953-970.

Kader, A.A. and Rahman, A.F.M.A. (2005). Testing the Weak-Form Efficiency of an Emerging Market: Evidence from the Dhaka Stock Exchange of Bangladesh. *AIUB Journal*, Vol. 4(2), August.

Kai, Fu and Wenhua, Xu (1997). Training neural network with genetic algorithms for forecasting the stock price index. 1997 IEEE International Conference on Intelligent Processing Systems (Cat. No.97TH8335).

Kang, S.H., Cho, G., and Yoon, S.M. (2009). Modeling Sudden Volatility Changes: Evidence from Japanese and Korean Stock Markets. *Physica A: Statistical Mechanics and its Applications*, 388(17), pp. 3543-3550. http://dx.doi.org/10.1016/j.physa.2009.05.028

Kassapi., S. (2016). VAR with first difference and level stationary variables?. Retrieved from: https://www.researchgate.net/post/VAR_with_first_difference_and_level_stationary_variables/57bd9577404854f99e25eb21/citation/download.

Kim, K.J. (2003). Financial time series forecasting using support vector machines. *Neurocomputing*, 55, pp. 307-319.

Kwiatowski et. al (1992). Testing the null hypothesis of stationarity against the alternative of a unit root. *Journal of Econometrics*, 54, pp. 159-178.

Leeflang, P.S.H. and Dick, R.W. (2000). Building Models for Marketing Decisions. *International Journal of Research in Marketing*, 17 (2), pp. 105–126.

Levine, R. and Zervos, S. (1998). Stock Markets, Banks, and Economic Growth. *The American Economic Review*, 88(3), pp. 537-558.

Liu, L.M. and Hudak, G.B. (1986). The SCA Statistical System, Reference Manual for Forecasting and Time Series Analysis, Version III, Lincoln Center.

Lutkepohl, H. (1991). Introduction to multiple time series analysis. Berlin: Springer-Verlag.

Lutkepohl, H. (1999). Vector autoregressions (unpublished manuscript, InstitutfürStatistik und Ökonometrie, Humboldt-Universitatzu, Berlin).

MacKinnon, J.G. (1996). Numerical Distribution Functions for Unit Root and Cointegration Tests. *Journal of Applied Econometrics*, 11, pp. 601-618.

Mills, T.C. (1999). The econometric modeling of financial time series (2nd edition). Cambridge: Cambridge University Press.

Mobarek, A. and Keasey, K. (2000). Weak-form market efficiency of an emerging market: Evidence from Dhaka Stock Market of Bangladesh. Paper presented at the ENBS Conference held on Osio, May. Available at http://www.bath.ac.uk/Centres/CDS/enbs-papers/mobarek_new.html (Accessed on March 22, 2018).

Müller, K., Smola, A., Rätsch, G., Schölkopf, B., Kohlmorgen, J. and Vapnik, V. (1997). Predicting time series with support vector machines. *Artificial Neural Networks*, 1327, pp. 999-1004.

Murray, M. (1994). A Drunk and Her Dog: An Illustration of Cointegration and Error Correction. *American Statistician - AMER STATIST*. 48, pp. 37-39. 10.1080/00031305.1994.10476017.

Nelson, C.R. (1976). Inflation and Rates of Return on Common Stocks. *Journal of Finance*, 31(2), pp. 471-483.

Ng, S. and Perron, P. (2001). Lag length selection and the construction of unit root tests with good size and power. *Econometrica*, 69(6), pp. 1519-1554.

Ojah, K., and Karemera, D. (1999). Random walks and Market efficiency Tests of Latin American Emerging Equity Markets: A Revisit. *The Financial Review*, 34(2), pp. 57-72.

Onwukwe, C.E., Samson, T.K. and Lipcsey, Z. (2014). Modelling and Forecasting Daily Returns Volatility of Nigerian Banks Stocks. *European Scientific Journal*, 10(15).

Pankratz, A. (1991). Forecasting with Dynamic Regression models, Wiley Inter Science.

Phillips, P.C.B. and Perron, P. (1988). Testing for a unit root in time series regression. *Biometrika*, 75, pp. 335-346.

Poshakwale, S. (1996). Evidence on the Weak-form efficiency and the day of the week effect in the Indian Stock Market. *Finance India*, 10(3), pp. 605-616.

Pu, Shen (2000). The P/E ratio and stock market performance. *Economic Review*, Federal Reserve Bank of Kansas City, issue Q IV, pp. 23-36.

Samsudin, R., Shabri, A. and Saad, P. (2010). A Comparison of Time Series Forecasting using Support Vector Machine and Artificial Neural Network Model. *Journal of Applied Sciences*, 10, pp. 950-958.

Ray, S. (2012). Foreign Exchange Reserve and Its Impact on Stock Market Capitalization: Evidence from India, *Research on Humanities and Social Sciences*, 2(2), pp. 46-60.

Reyes, M.G. (2001). Asymmetric Volatility spillover in the Tokyo Stock Exchange. *Journal of Economics and Finance,* 25(2), pp. 206-213. http://dx.doi.org/10.1007/BF02744523

Rob J., Hyndman et al. (2020). Forecasting Functions for Time Series and Linear Models, R Package 'forecast', Version 8.13. URL https://pkg.robjhyndman.com/forecast/

Roux, F.J.P. and Gilbertson, D.P. (1978). The behavior of share prices on the Johannesburg Stock Exchange. *Journal of Business Finance and Accounting*, 5(2), pp. 223-232.

Roy, M.L. and Ashrafuzzaman, M. (2015). EMH, Earning Multiples and Common Stock Valuation: The Case of Dhaka Stock Exchange. *International Review of Business Research Papers*, 11(2), pp. 62-76.

Sattayatham, P., Sopipan, N., and Premanode, C. (2012). Forecasting the Stock Exchange of Thailand using day of the week effect and Markov Regime Switching GARCH. *American Journal of Economics and Business Administration*, 4(1), pp. 84-93.

Schwarz, G.E. (1978). Estimating the dimension of a model. *Annals of Statistics,* 6 (2), pp. 461–464.

Shah, D., Isah, H., and Zulkernine, F. (2019). Stock Market Analysis: A Review and Taxonomy of Prediction Techniques, *International Journal of Financial Studies*, *7*(2), p. 26.

Sharma, P. and Vipul, V. (2015). Forecasting stock index volatility with GARCH models: International evidence. *Studies in Economics and Finance*. Forthcoming. 10.1108/SEF-11-2014-0212.

Shen,W., Guo, X., Wu, C. and Wu, D. (2011). Forecasting stock indices using radial basis function neural networks optimized by artificial fish swarm algorithm. *Knowledge-Based Systems*, 24(3), pp. 378–385.

Sims, C.A. (1980). Macroeconomics and reality. *Econometrica*, 48, pp. 1-48.

Smola, A.J. and Schölkopf, B. (2004). A tutorial on support vector regression. *Statistics and Computing,* 14(3), pp. 199-222.

Startz (2009). Non-normality and collinearity are NOT problems! EViews user forum. Retrieved from http://forums.eviews.com/viewtopic.php?t=512

Tay, F.E.H. and Cao, L. (2001). Application of support vector machines in financial time series forecasting. *Omega*, 29(4), pp. 309-317.

Thissen, U., Van Brakel, R., de Weijer, A.P., Melssen, W.J. and Buydens, L.M.C. (2003). Using support vector machines for time series prediction. *Chemometrics and Intelligent Laboratory Systems,* 69(1-2), pp. 35-49.

Tsay, R.S. (2001). Analysis of financial time series, Berlin: Springer-Verlag, pp. 312-318.

Uddin, M.G.S. and Alam, M.M. (2007). The Impacts of Interest Rate on Stock Market: Empirical Evidence from Dhaka Stock Exchange. *South Asian Journal of Management and Sciences*, 1(2), pp. 123-132.

Deneshkumar, V. and Sentham, K. (2011). Outliers in time series data. *International Journal of Agricultural and Statistical Sciences*, 7, pp. 685-691.

Vapnik, V. (1979). Estimation of Dependences Based on Empirical Data (in Russian). Moscow, Nauca.

Vapnik, V. (1998). Statistical learning theory. New York, Wiley.

Vapnik, V. (1999). An overview of statistical learning theory. IEEE Transactions on Neural Networks, 10, pp. 988-999.

Vee, D.N.C. and Gonpot, P.N.S. (2011). Forecasting USD/MUR Exchange Rate using a GARCH (1,1) model with GED and Student's-t errors. *University of Mauritius Research Journal*, 17.

Waggoner, D.F. and Zha, T. (1999). Conditional forecasts in dynamic multivariate models. *Review of Economics and Statistics*, 81(4), pp. 639-651.

Wang, W.C., Chau, K.W., Chen, C.T. and Qiu, L. (2009). A comparison of performance of several artificial intelligence methods for forecasting monthly discharge time series. *J. Hydrol.*, 374, pp. 294-306.

Watson, M. (1994). Vector autoregressions and cointegration. In R. F. Engle and D. MacFadden (Eds.), Handbook of econometrics. Amsterdam: Elsevier Science Ltd.

World Bank WDI (2017). World Development Indicators. Retrieve from https://datacatalog.worldbank.org/dataset/world-development-indicators

Appendix-A

Variables and Time Series Data Sets

Table A: Annual Data of STR, TEC and IMC (US$) (During 1990 to 2012)

Year	STR	TEC	IMC (US$)
1990	1.505646173	134	321000000
1991	1.016949153	138	269000000
1992	3.773584906	145	314000000
1993	3.911342894	153	453000000
1994	14.23819029	170	1050000000
1995	13.23283082	183	1338000000
1996	24.52029207	186	4551000000
1997	12.6478318	202	1537000000
1998	61.37689615	208	1034000000
1999	83.02607456	211	865403300
2000	74.83450072	221	1185950000
2001	63.6006711	230	1144560000
2002	56.94912064	239	1192930000
2003	23.24654283	247	1621510000
2004	36.05474548	250	3316980000
2005	31.49780082	262	3035400000
2006	28.37701598	269	3610260000
2007	92.28960666	278	6793230000
2008	137.2587394	290	6670562037
2009	212.561601	302	7067627057
2010	129.163043	209	15683000000
2011	92.5753602	216	23546000000
2012	61.17633323	229	17479000000

Data Source: World Bank Data Indicator, Country Bangladesh
http://data.worldbank.org/country/bangladesh

Table B: Annual Data of DGI, GDP, GNI and GS of Bangladesh (During 2005 to 2012)

Year	DGI	GDP (US$)	GNI (US$)	GS (US$)
2005	20976.18	60277560976	63355284553	18739677715
2006	18194.88	61901116736	65952263252	21361399703
2007	25444.77	68415421373	73522980017	24395379769
2008	34537.74	79554350678	86607185541	29237022626
2009	34224.34	89359767442	97484941860	34518222526
2010	73842.72	1.00357E+11	1.09695E+11	38584264100
2011	74368.28	1.11879E+11	1.22062E+11	40820088900
2012	54339.48	1.16355E+11	1.27672E+11	46279032616

Data Source: World Bank Data Indicator, Country Bangladesh
http://data.worldbank.org/country/bangladesh

Table C: Annual Data of GI, DIR (%) and GFI of Bangladesh (During 2005 to 2012)

Year	GI (%)	DIR (%)	GFI (US$)
2005	7.0466182	8.0925	813321971.9
2006	6.7652612	9.1125	697206284.1
2007	9.106985	9.1759	652818718.9
2008	8.9019449	9.6533	1009623164
2009	5.4234724	8.2050	732809635.6
2010	8.1266764	7.1425	918172637.9
2011	10.704805	10.015	1137916361
2012	6.2181824	11.685	1178439622

Data Source: World Bank Data Indicator, Country Bangladesh
http://data.worldbank.org/country/bangladesh

Table D: Monthly Data of Capital in Taka (mn), Volume, Value in Taka (mn), Trade and DGI from DSE (June 2004 to July 2013)

Month	CAPITAL	VOLUME	VALUE	TRADE	DGI
2004M06	131848.179	3305352.200	178.485	9667.160	1273.014
2004M07	132126.825	1777211.217	119.702	6426.520	1269.784
2004M08	146526.932	3360264.000	245.583	10483.090	1418.756
2004M09	139063.027	4639028.333	277.303	11873.000	1592.645
2004M10	168381.123	1919888.750	253.457	9196.460	1705.919
2004M11	183984.900	2887991.737	387.353	11936.890	1819.067
2004M12	197093.605	4051619.652	414.260	13356.170	1939.044
2005M01	219500.952	2559046.650	243.097	8830.450	1869.809
2005M02	220825.524	3260732.647	246.082	11093.310	1814.054
2005M03	233324.102	5409300.880	244.589	9692.710	1936.914
2005M04	216693.745	4335016.764	446.755	13699.160	1895.202
2005M05	202698.332	4739733.083	273.361	10551.910	1895.202
2005M06	218384.757	4072147.609	236.833	9044.440	1771.189
2005M07	212603.704	3545199.400	233.444	8133.880	1599.919
2005M08	207936.348	1917914.750	206.843	7338.480	1692.452
2005M09	220546.846	3033413.160	173.585	10201.500	1619.115
2005M10	224841.440	4245388.500	265.222	8170.570	1564.764
2005M11	229142.838	2961041.190	203.480	11382.270	1647.065
2005M12	225163.645	3040945.000	282.937	8441.790	1670.493
2006M01	226652.537	1625757.789	193.697	7771.770	1659.911
2006M02	1600.037	1937216.846	141.870	9347.250	1701.267
2006M03	225016.370	2247798.350	159.192	11047.670	1649.958
2006M04	206302.258	2310462.133	198.975	10197.460	1546.419
2006M05	208483.254	2509405.810	197.094	10826.430	1391.362
2006M06	211012.816	2142168.316	157.921	8756.210	1379.536
2006M07	224078.642	2356187.952	235.112	12712.570	1326.537
2006M08	260009.524	6406092.476	560.745	24051.140	1364.494
2006M09	281654.839	4932078.889	553.314	21486.110	1531.664
2006M10	278914.418	3194915.313	289.420	12288.630	1589.158
2006M11	298437.693	4873717.000	351.017	14577.160	1538.310
2006M12	311744.859	9354059.833	439.110	16896.110	1516.263
2007M01	336622.925	18180691.200	809.411	25540.000	1582.640
2007M02	381081.764	18824181.890	1109.610	32838.580	1684.278
2007M03	380915.870	8991656.300	642.009	23163.900	1831.499
2007M04	388627.391	5143141.571	539.456	17854.430	1757.588
2007M05	420342.049	8756565.476	1259.286	33718.760	1712.795
2007M06	454371.234	10254290.700	1545.512	37188.600	1867.256
2007M07	510312.442	15820481.410	1940.849	44336.640	2050.596
2007M08	544948.857	9165436.900	1386.034	33370.100	2289.590

Month	CAPITAL	VOLUME	VALUE	TRADE	DGI
2007M09	598043.875	11984598.850	1553.187	34640.200	2349.564
2007M10	673609.838	18305745.940	2297.278	51508.890	2551.083
2007M11	726205.347	11983370.050	2066.441	48872.850	2801.128
2007M12	728976.873	5801660.813	1244.309	31261.500	2966.752
2008M01	775458.671	8996626.045	1520.326	36198.680	2942.323
2008M02	794278.686	16471068.790	2102.802	48579.630	2931.355
2008M03	815593.583	19032819.100	2831.004	67543.700	2915.230
2008M04	2981.711	25924834.480	3299.439	76439.620	3073.242
2008M05	838914.461	23278932.630	3681.584	75463.740	3085.751
2008M06	861494.007	20445684.820	3179.769	71568.050	3068.263
2008M07	891246.177	20173895.730	2957.643	65942.140	2904.328
2008M08	966123.345	15103696.110	2400.689	55315.210	2696.164
2008M09	944008.577	25487563.050	3448.745	72524.420	2853.212
2008M10	998069.126	28672462.050	4162.858	83887.890	2861.964
2008M11	1003868.939	14884119.950	2075.696	54527.570	2628.830
2008M12	962973.469	14636939.570	2132.834	53947.070	2577.077
2009M01	982017.746	28454176.650	3293.248	78343.200	2706.424
2009M02	1028003.023	23038192.700	2867.807	80642.500	2580.995
2009M03	999480.114	30649585.810	4549.531	110155.710	2589.342
2009M04	1016767.192	23824510.190	4398.613	102291.950	2520.920
2009M05	1023753.637	26535395.000	4910.794	98837.190	2555.131
2009M06	1034360.926	34068045.860	6853.608	130856.360	2535.580
2009M07	1138171.228	34745811.430	5890.392	114231.710	2795.933
2009M08	1244582.583	39995250.760	6395.897	123035.100	2911.521
2009M09	1294553.126	31448941.060	5166.482	98230.060	2982.623
2009M10	1326761.602	55774547.860	10024.096	161178.570	2994.208
2009M11	1417911.337	32855451.950	8990.878	144107.350	3276.438
2009M12	1622333.281	29320989.900	8963.163	143151.850	3775.220
2010M01	1866613.237	48604501.810	12517.590	186657.670	4415.349
2010M02	2044037.591	50191352.580	13156.953	180443.740	4941.337
2010M03	2292537.689	33999742.910	7959.817	112123.770	5612.953
2010M04	2259027.054	35031457.350	9565.159	132715.200	5527.412
2010M05	2306251.577	48285036.190	18392.302	225664.290	5574.668
2010M06	2429873.623	54724528.550	17624.256	207232.270	5819.689
2010M07	2667689.104	61222318.370	16987.856	211752.320	6207.813
2010M08	2790637.911	81519611.770	17929.003	228425.320	6354.707
2010M09	2961409.716	89046121.720	17364.575	226097.890	6634.050
2010M10	3060230.461	113267378.300	23400.448	299698.650	6897.628
2010M11	3265863.379	117543826.300	24827.221	316926.470	7548.645
2010M12	3492395.858	108355932.700	18436.861	253477.480	8308.462
2011M01	3512211.567	72132469.900	9348.462	143008.000	8339.505

Month	CAPITAL	VOLUME	VALUE	TRADE	DGI
2011M02	3216375.631	60022099.060	6758.066	132856.180	7415.153
2011M03	2845752.815	89449473.000	9869.324	188883.770	6321.154
2011M04	2771706.023	72416180.210	8225.493	166768.260	6142.291
2011M05	2826324.767	49303059.760	4407.916	118042.620	6272.803
2011M06	2630077.983	69060468.270	6120.476	154704.640	5644.179
2011M07	2714956.488	172337698.300	14908.649	281857.800	5845.086
2011M08	3020613.217	64914582.060	5214.626	111525.590	6481.582
2011M09	2927389.226	40971534.600	3528.624	86628.850	6176.487
2011M10	2674134.612	44282863.190	3332.889	97337.570	5468.238
2011M11	2583009.962	60358116.290	4089.109	121158.410	5178.807
2011M12	2546477.495	67361689.630	3238.411	103556.470	5082.998
2012M01	2508342.772	81116705.550	4035.023	118801.320	4921.320
2012M02	2242891.131	72616085.050	3012.827	104915.790	4199.880
2012M03	2403521.054	87702223.600	4206.155	116317.800	4607.035
2012M04	2693699.637	133854797.700	8027.393	171109.140	5245.798
2012M05	2596326.692	60454484.430	3186.580	82662.900	4933.975
2012M06	2475126.641	42074896.200	1968.187	66158.650	4562.257
2012M07	2361270.115	44488765.480	1931.624	67901.000	4185.059
2012M08	2377459.504	92173847.640	4091.186	110253.430	4228.694
2012M09	2523704.976	198440255.900	8934.053	194549.810	4547.435
2012M10	2528586.158	141151170.400	5605.015	127125.260	4532.278
2012M11	2402904.899	74853293.600	2769.035	78410.700	4255.544
2012M12	2356225.191	60191147.740	2143.743	71438.320	4120.210
2013M01	2373095.049	51792554.650	1693.817	54691.740	4156.421
2013M02	2424548.422	93797194.790	3795.628	109108.370	4263.981
2013M03	2278420.959	46225488.110	1850.316	68280.330	3883.516
2013M04	2193355.598	49571662.570	1536.390	63301.240	3673.183
2013M05	2290067.745	91250697.200	3000.793	98178.500	3893.764
2013M06	2502624.507	158504876.300	6618.310	152870.700	4335.967
2013M07	2585840.744	125662029.500	6956.657	138560.360	4508.987

Data Source: Dhaka Stock Exchange (DSE) Limited, Bangladesh
https://www.dsebd.org/recent_market_information.php

Table E: Monthly Data of DSEX, DSES and DSE30 Indices from DSE (January 2014 to December 2018)

Months	DSEX	DSES	DSE30
2014M01	4534.7150	968.5243	1593.6236
2014M02	4765.1739	989.5993	1680.1419
2014M03	4599.6916	996.8329	1652.5428
2014M04	4611.0170	1022.1466	1684.2775
2014M05	4449.1328	991.6566	1620.1377
2014M06	4394.9264	1006.3490	1620.8598
2014M07	4392.7368	996.6344	1608.1169
2014M08	4534.4558	1050.5842	1698.0559
2014M09	4796.5921	1126.8850	1841.3879
2014M10	5205.1500	1222.3782	1964.2716
2014M11	4915.3508	1151.9898	1822.4726
2014M12	4888.2086	1148.1129	1808.9299
2015M01	4871.2433	1155.5376	1809.5250
2015M02	4752.2522	1126.9201	1768.1670
2015M03	4576.6667	1103.0649	1719.9383
2015M04	4297.6185	1050.6796	1642.0212
2015M05	4369.6454	1059.8557	1651.0242
2015M06	4527.0166	1104.2424	1744.1395
2015M07	4675.9546	1153.6251	1828.8925
2015M08	4816.8986	1188.4670	1857.8395
2015M09	4799.5737	1178.8838	1834.6762
2015M10	4708.3103	1130.6826	1786.5449
2015M11	4507.9922	1085.7837	1711.6368
2015M12	4587.6479	1104.7772	1743.1065
2016M01	4646.1827	1117.4318	1754.0939
2016M02	4572.3402	1115.4651	1751.7295
2016M03	4420.9927	1072.8535	1686.9026
2016M04	4365.5370	1058.9198	1664.7165
2016M05	4337.3820	1064.8887	1686.2459
2016M06	4412.6216	1086.0562	1734.8253
2016M07	4542.5420	1114.3913	1775.7318
2016M08	4563.1358	1115.8036	1768.5177
2016M09	4632.3724	1115.0661	1765.2792
2016M10	4684.3780	1118.4207	1760.9363
2016M11	4708.4237	1124.2592	1759.1902

Months	DSEX	DSES	DSE30
2016M12	4916.9438	1166.6513	1795.4275
2017M01	5398.3663	1254.5622	1940.6421
2017M02	5531.5736	1291.6467	2008.4692
2017M03	5674.2655	1304.8583	2054.8085
2017M04	5608.3453	1289.3139	2075.6724
2017M05	5442.9835	1262.5482	2010.1783
2017M06	5503.1337	1270.3416	2043.4328
2017M07	5793.4550	1315.6979	2121.4829
2017M08	5899.2017	1310.3855	2119.2675
2017M09	6145.6118	1361.0503	2193.0482
2017M10	6062.6914	1330.5491	2185.8816
2017M11	6229.9395	1362.4730	2252.6526
2017M12	6224.7637	1378.2451	2253.5448
2018M01	6181.0446	1404.9937	2270.4937
2018M02	5945.5246	1385.1607	2198.8766
2018M03	5694.0549	1343.5353	2112.5254
2018M04	5809.7397	1351.4323	2179.7238
2018M05	5503.0463	1282.2581	2045.1814
2018M06	5380.7249	1247.8233	1968.1356
2018M07	5329.5612	1261.0455	1902.1184
2018M08	5468.3508	1255.3191	1921.1658
2018M09	5477.9521	1260.5224	1918.9291
2018M10	5356.5061	1241.2446	1891.8016
2018M11	5261.2461	1214.5108	1858.5156
2018M12	5294.2992	1218.1788	1854.1851

Data Source: Dhaka Stock Exchange (DSE) Limited, Bangladesh
https://www.dsebd.org/recent_market_information.php

Appendix-B

Actual, Fitted, and Residual Series of Proposed Models

Table F: Actual, Fitted and Residuals obtained from ARIMA(1,0,0) with EGARCH(1,1,2) model of DSEX (January 2014 to December 2018)

Month	Actual	Fitted	Residual	Standardized Residual
Jan-14	4534.715	NA	NA	NA
Feb-14	4765.174	4554.324	210.8504	1.302
Mar-14	4599.692	4775.376	-175.685	-0.828
Apr-14	4611.017	4616.648	-5.6312	-0.045
May-14	4449.133	4627.511	-178.379	-1.314
Jun-14	4394.926	4472.234	-77.308	-0.602
Jul-14	4392.737	4420.24	-27.5035	-0.225
Aug-14	4534.456	4418.14	116.3157	0.872
Sep-14	4796.592	4554.075	242.5172	1.272
Oct-14	5205.15	4805.512	399.6377	1.897
Nov-14	4915.351	5197.395	-282.044	-1.185
Dec-14	4888.209	4919.424	-31.2153	-0.288
Jan-15	4871.243	4893.39	-22.1463	-0.189
Feb-15	4752.252	4877.117	-124.864	-0.929
Mar-15	4576.667	4762.982	-186.315	-1.402
Apr-15	4297.619	4594.563	-296.945	-2.564
May-15	4369.645	4326.904	42.7414	0.488
Jun-15	4527.017	4395.991	131.0254	1.203
Jul-15	4675.955	4546.939	129.0152	0.777

Month	Actual	Fitted	Residual	Standardized Residual
Aug-15	4816.899	4689.799	127.1	0.820
Sep-15	4799.574	4824.99	-25.4163	-0.150
Oct-15	4708.31	4808.372	-100.062	-0.729
Nov-15	4507.992	4720.834	-212.841	-1.535
Dec-15	4587.648	4528.691	58.9565	0.497
Jan-16	4646.183	4605.096	41.0867	0.288
Feb-16	4572.34	4661.242	-88.9016	-0.592
Mar-16	4420.993	4590.413	-169.42	-1.215
Apr-16	4365.537	4445.243	-79.7057	-0.633
May-16	4337.382	4392.051	-54.6685	-0.450
Jun-16	4412.622	4365.045	47.577	0.370
Jul-16	4542.542	4437.213	105.3287	0.662
Aug-16	4563.136	4561.831	1.3047	0.007
Sep-16	4632.372	4581.584	50.788	0.331
Oct-16	4684.378	4647.995	36.3829	0.201
Nov-16	4708.424	4697.878	10.5455	0.060
Dec-16	4916.944	4720.942	196.0013	1.137
Jan-17	5398.366	4920.952	477.4144	1.999
Feb-17	5531.574	5382.726	148.8481	0.545
Mar-17	5674.266	5510.496	163.7695	1.051
Apr-17	5608.345	5647.364	-39.0188	-0.202
May-17	5442.984	5584.134	-141.151	-1.023
Jun-17	5503.134	5425.522	77.6121	0.584
Jul-17	5793.455	5483.217	310.2381	1.898
Aug-17	5899.202	5761.689	137.5128	0.554
Sep-17	6145.612	5863.12	282.4922	1.860

Month	**Actual**	**Fitted**	**Residual**	**Standardized Residual**
Oct-17	6062.691	6099.473	-36.7812	-0.155
Nov-17	6229.94	6019.937	210.0029	1.673
Dec-17	6224.764	6180.359	44.4051	0.192
Jan-18	6181.045	6175.394	5.6506	0.040
Feb-18	5945.525	6133.459	-187.935	-1.246
Mar-18	5694.055	5907.552	-213.497	-1.560
Apr-18	5809.74	5666.346	143.394	1.280
May-18	5503.046	5777.309	-274.263	-1.602
Jun-18	5380.725	5483.133	-102.408	-1.023
Jul-18	5329.561	5365.804	-36.2428	-0.372
Aug-18	5468.351	5316.729	151.6221	1.372
Sep-18	5477.952	5449.854	28.0986	0.145
Oct-18	5356.506	5459.063	-102.557	-0.756
Nov-18	5261.246	5342.574	-81.3276	-0.608
Dec-18	5294.299	5251.202	43.0975	0.328

Table G: Actual, Fitted and Residuals obtained from ANN (MLP 3-7-1 net) model of DSES (January 2014 to December 2018)

Month	Actual	Fitted	Residual	Standardized Residual
Jan-14	968.52	NA	NA	NA
Feb-14	989.6	NA	NA	NA
Mar-14	996.83	1025.08	-28.250	-0.9465
Apr-14	1022.15	1021.39	0.760	-0.0354
May-14	991.66	1051.82	-60.160	-1.9487
Jun-14	1006.35	996.07	10.280	0.2636
Jul-14	996.63	1032.99	-36.360	-1.2012
Aug-14	1050.58	1010.95	39.630	1.1854
Sep-14	1126.88	1092.4	34.480	1.0236
Oct-14	1222.38	1158.21	64.170	1.9561
Nov-14	1151.99	1224.39	-72.400	-2.3331
Dec-14	1148.11	1121.24	26.870	0.7846
Jan-15	1155.54	1144.85	10.690	0.2765
Feb-15	1126.92	1156.48	-29.560	-0.9876
Mar-15	1103.06	1111.88	-8.820	-0.3363
Apr-15	1050.68	1090.36	-39.680	-1.3055
May-15	1059.86	1026.72	33.140	0.9816
Jun-15	1104.24	1071.35	32.890	0.9737
Jul-15	1153.63	1128.08	25.550	0.7432
Aug-15	1188.47	1168.68	19.790	0.5623
Sep-15	1178.88	1194.29	-15.410	-0.5432
Oct-15	1130.68	1174.76	-44.080	-1.4436
Nov-15	1085.78	1106.36	-20.580	-0.7056
Dec-15	1104.78	1061.94	42.840	1.2862
Jan-16	1117.43	1115.41	2.020	0.0042
Feb-16	1115.47	1123.28	-7.810	-0.3045

Month	Actual	Fitted	Residual	Standardized Residual
Mar-16	1072.85	1114.04	-41.190	-1.3529
Apr-16	1058.92	1051.27	7.650	0.1810
May-16	1064.89	1056.11	8.780	0.2165
Jun-16	1086.06	1073.55	12.510	0.3336
Jul-16	1114.39	1100.66	13.730	0.3720
Aug-16	1115.8	1128.39	-12.590	-0.4547
Sep-16	1115.07	1116.1	-1.030	-0.0916
Oct-16	1118.42	1114.3	4.120	0.0701
Nov-16	1124.26	1119.52	4.740	0.0896
Dec-16	1166.65	1126.17	40.480	1.2121
Jan-17	1254.56	1177.23	77.330	2.3694
Feb-17	1291.65	1248.98	42.670	1.2809
Mar-17	1304.86	1290.46	14.400	0.3930
Apr-17	1289.31	1307.05	-17.740	-0.6164
May-17	1262.55	1295.2	-32.650	-1.0847
Jun-17	1270.34	1266.81	3.530	0.0516
Jul-17	1315.7	1273.76	41.940	1.2579
Aug-17	1310.39	1310.58	-0.190	-0.0652
Sep-17	1361.05	1314.83	46.220	1.3923
Oct-17	1330.55	1346.32	-15.770	-0.5545
Nov-17	1362.47	1335.15	27.320	0.7988
Dec-17	1378.25	1351.46	26.790	0.7821
Jan-18	1404.99	1365.01	39.980	1.1964
Feb-18	1385.16	1378.85	6.310	0.1389
Mar-18	1343.54	1373.16	-29.620	-0.9895
Apr-18	1351.43	1346.21	5.220	0.1047
May-18	1282.26	1347.63	-65.370	-2.1123
Jun-18	1247.82	1286.91	-39.090	-1.2869

Month	Actual	Fitted	Residual	Standardized Residual
Jul-18	1261.05	1249.58	11.470	0.3010
Aug-18	1255.32	1263.8	-8.480	-0.3256
Sep-18	1260.52	1258.71	1.810	-0.0024
Oct-18	1241.24	1263.8	-22.560	-0.7678
Nov-18	1214.51	1242.88	-28.370	-0.9503
Dec-18	1218.18	1211.1	7.080	0.1631

Table H: Actual, Fitted and Residuals obtained from ARIMA(1,0,1) with ARCH(2) model of DSE30 (January 2014 to December 2018)

Month	Actual	Fitted	Residual	Standardized Residual
Jan-14	1593.624	NA	NA	NA
Feb-14	1680.142	1695.176	-15.0344	-0.1242
Mar-14	1652.543	1748.466	-95.9229	-0.7122
Apr-14	1684.278	1702.938	-18.6603	-0.1487
May-14	1620.138	1750.16	-130.022	-1.0905
Jun-14	1620.86	1669.305	-48.4456	-0.4864
Jul-14	1608.117	1696.372	-88.2554	-1.1594
Aug-14	1698.056	1674.55	23.5064	0.1927
Sep-14	1841.388	1773.47	67.9177	0.5484
Oct-14	1964.272	1887.578	76.694	0.5586
Nov-14	1822.473	1975.865	-153.393	-1.2850
Dec-14	1808.93	1802.36	6.5703	1.2436
Jan-15	1809.525	1845.035	-35.5102	-0.9065
Feb-15	1768.167	1831.746	-63.5789	-0.4307
Mar-15	1719.938	1793.853	-73.9148	-0.5415
Apr-15	1642.021	1756.958	-114.937	-0.9362
May-15	1651.024	1689.431	-38.4071	-0.4280
Jun-15	1744.14	1720.612	23.5278	0.2380
Jul-15	1828.893	1805.515	23.3778	0.1612
Aug-15	1857.84	1864.387	-6.5472	-0.0443
Sep-15	1834.676	1874.766	-40.0897	-0.2688
Oct-15	1786.545	1847.74	-61.1949	-0.4170
Nov-15	1711.637	1807.406	-95.7691	-0.7033
Dec-15	1743.107	1744.071	-0.964	-0.0086
Jan-16	1754.094	1796.821	-42.727	-0.3548
Feb-16	1751.73	1790.86	-39.1302	-0.2674

Month	Actual	Fitted	Residual	Standardized Residual
Mar-16	1686.903	1790.387	-103.485	-0.7320
Apr-16	1664.717	1724.363	-59.646	-0.5143
May-16	1686.246	1723.214	-36.9685	-0.3611
Jun-16	1734.825	1745.567	-10.7412	-0.0789
Jul-16	1775.732	1787.88	-12.1481	-0.0829
Aug-16	1768.518	1815.86	-47.3426	-0.3153
Sep-16	1765.279	1799.384	-34.105	-0.2355
Oct-16	1760.936	1801.443	-40.5071	-0.2873
Nov-16	1759.19	1796.339	-37.1492	-0.2595
Dec-16	1795.428	1796.219	-0.7915	-0.0056
Jan-17	1940.642	1833.251	107.3911	0.7320
Feb-17	2008.469	1969.434	39.0352	0.3274
Mar-17	2054.809	1994.328	60.4803	0.5703
Apr-17	2075.672	2033.527	42.1453	0.3084
May-17	2010.178	2042.061	-31.8828	-0.2366
Jun-17	2043.433	1972.423	71.0103	0.4970
Jul-17	2121.483	2029.048	92.4353	0.6868
Aug-17	2119.268	2090.285	28.9821	0.2658
Sep-17	2193.048	2068.082	124.9661	1.0398
Oct-17	2185.882	2150.632	35.2501	0.3451
Nov-17	2252.653	2116.434	136.2187	1.5449
Dec-17	2253.545	2195.733	57.8116	0.6507
Jan-18	2270.494	2170.821	99.673	1.6681
Feb-18	2198.877	2196.236	2.6411	0.0237
Mar-18	2112.525	2114.849	-2.3238	-0.0198
Apr-18	2179.724	2053.201	126.5233	0.8388
May-18	2045.181	2141.875	-96.694	-0.9263
Jun-18	1968.136	1975.652	-7.516	-0.1999

Month	Actual	Fitted	Residual	Standardized Residual
Jul-18	1902.118	1951.129	-49.0106	-0.4099
Aug-18	1921.166	1891.721	29.4448	0.2035
Sep-18	1918.929	1930.511	-11.5821	-0.0820
Oct-18	1891.802	1915.596	-23.7947	-0.1608
Nov-18	1858.516	1892.76	-34.2445	-0.2297
Dec-18	1854.185	1866.217	-12.0315	-0.0823

Appendix-C

R Code

#Install Packages

```
install.packages("forecast")
install.packages("tseries")
install.packages("zoo ")
install.packages("parallel")
install.packages("car")
install.packages("moments")
install.packages("FinTS")
```

#Load packages

```
library(forecast)
library(tseries)
library(zoo)
library(parallel)
library(car)
library(moments)
library(FinTS)
```

#Read Data of DSE Market Capital

```
setwd("F:/phd_data")
getwd()
capital = read.csv("capital.csv", header = T)
capital
```

#Read Data of DSE General Index (DGI)

```
setwd("F:/phd_data")
getwd()
dgi = read.csv("dgi.csv", header = T)
dgi
```

#Read Data of DSE Market Value

```
setwd("F:/phd_data")
getwd()
value = read.csv("value.csv", header = T)
value
```

#Read Data of DSE Market Volume

```
setwd("F:/phd_data")
getwd()
volume = read.csv("volume.csv", header = T)
volume
```

#Read Data of DSE Trade

```
setwd("F:/phd_data")
getwd()
trade = read.csv("trade.csv", header = T)
trade
```

#Read Data of DSEX

```
setwd("F:/phd_data")
getwd()
dsex = read.csv("dsex.csv", header = T)
dsex
```

#Read Data of DSES

```
setwd("F:/phd_data")
getwd()
dses = read.csv("dses.csv", header = T)
dses
```

#Read Data of DSE30

```
setwd("F:/phd_data")
getwd()
dse30 = read.csv("dse30.csv", header = T)
dse30
```

#Auto ARIMA Estimation of DSE Market Capital

```
library(forecast)
setwd("F:/phd_data")
getwd()
capital=read.csv("capital.csv", header=T)
capital
y=auto.arima(capital)
y
```

#Auto ARIMA Estimation of DSE General Index (DGI)

```
library(forecast)
setwd("F:/phd_data")
getwd()
dgi=read.csv("dgi.csv", header=T)
dgi
y=auto.arima(dgi)
y
```

#Auto ARIMA Estimation of DSE Market Value

```
library(forecast)
setwd("F:/phd_data")
getwd()
value=read.csv("value.csv", header=T)
value
y=auto.arima(value)
value
```

#Auto ARIMA Estimation of DSE Market Volume

```
library(forecast)
setwd("F:/phd_data")
getwd()
volume=read.csv("volume.csv", header=T)
volume
y=auto.arima(volume)
volume
```

#Auto ARIMA Estimation of DSE Trade

```
library(forecast)
setwd("F:/phd_data")
getwd()
trade=read.csv("trade.csv", header=T)
trade
y=auto.arima(trade)
trade
```

#Auto ARIMA Estimation of DSEX

```
library(forecast)
setwd("F:/phd_data")
getwd()
dsex=read.csv("dsex.csv", header=T)
dsex
y=auto.arima(dsex)
y
```

#Auto ARIMA Estimation of DSES

```
library(forecast)
setwd("F:/phd_data")
getwd()
dses=read.csv("dses.csv", header=T)
dses
y=auto.arima(dses)
y
```

#Auto ARIMA Estimation of DSE30

```
library(forecast)
setwd("F:/phd_data")
getwd()
dse30=read.csv("dse30.csv", header=T)
dse30
y=auto.arima(dse30)
```

Appendix-D

Assumptions of the Estimated Models

Table I: Assumptions of the Estimated Models

Model Name	Estimation Methods	Assumptions
CD Functional Regression	Ordinary Least Squares (OLS)	• The coefficients and error terms of the regression model are linear. • The population mean of the error term is zero • The independent variables are uncorrelated with the error term. • The error term is not correlated with one another. • The error term has a constant variance (there is no heteroscedasticity). • There is no perfect linear relationship between independent variables and explanatory variables. • The error term is normally distributed (optional).
Multiple Linear Regression	OLS	Same as OLS assumptions
ARIMA	OLS	Same as OLS assumptions
VAR	OLS	Same as OLS assumptions, but the error terms are multivariate normally distributed.

Model Name	Estimation Methods	Assumptions
ARIMA with GARCH Family	ML - ARCH (Marquardt)	We must make certain assumptions in order to use ML - ARCH (Marquardt), which are typically referred to as the *i.i.d.* assumption. According to these assumptions, • The data must be distributed independently. • The data must be identically distributed. • The error term must be normally distributed.
ANN	ANN	• As information enters the network through its input layer, it passes through its output layer. • The input layer is the only way to provide input to the neural network. • There is no way to add information to the hidden layer of the neural network.
SVM regression	SVM	SVMs are linear classifiers under two assumptions: • The margin should be large. • Data points that are more likely to be incorrectly classified are support vectors, which are the most useful.