Portfolio Optimization Using ARIMA – Global Minimum Variance Approach

MSc Research Project
FinTech

Kamran Raiysat
Student ID: x19102429

School of Computing
National College of Ireland

Supervisor: Noel Cosgrave

National College of Ireland

MSc Project Submission Sheet

School of Computing

Student Name:	Kamran Raiysat		
Student ID:	x19102429		
Programme:	MSc in FinTech	**Year:**	2019-2020
Module:	Research Project		
Supervisor:	Noel Cosgrave		
Submission Due Date:	August 17, 2020		
Project Title:	Portfolio Optimization Using ARIMA – Global Minimum Variance Approach		
Word Count:	7338	**Page Count**	22

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.
ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature: Kamran Raiysat

Date: 27/09/2020

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST

Attach a completed copy of this sheet to each project (including multiple copies)	□
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	□
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	□

Assignments that are submitted to the Programme Coordinator Office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

Portfolio Optimization Using ARIMA-GMV Approach

Kamran Raiysat
Student ID: x19102429

Abstract

The stock price prediction and portfolio optimization have been in the academic sphere over a long period. Several techniques have been studied for this purpose but ARIMA and GMV are the most widely used statistical methods. Advanced machine learning has made statistical techniques more powerful. The study tries to find the performance of the ARIMA and GMV hybrid approach against the GMV approach on the KSE100 index. The purpose of the selection of the KSE100 index is that very limited studies have been conducted on Pakistan Stock Exchanges using advanced machine learning statistical approaches. The data is collected for six most capitalized stocks listed at KSE100 over five years using weekly ending prices. The data is prepared and split into 80% training and 20% test data. The Auto-ARIMA function is used to forecast the future expected values of each stock and KSE100 based on the training data. The validation dataset is created by combining training data and future expected values. Portfolio weights are calculated using GMV on training data and ARIMA-GMV using validation data. The results indicate that the ARIMA – GMV approach performs better than the traditional GMV approach in portfolio optimization. The ARIMA – GMV mainly depends on the performance of the ARIMA for the accuracy of the future expected value. It is also suggested to use non – linear techniques to optimize the portfolio using GMV to build an accurate model in a highly volatile market across borders.

Keywords: *ARIMA, GMV, Stock Markets, Forecast, Portfolio optimization*

1. Introduction

The importance of the stock markets for any country cannot be ignored because these markets play a crucial role in economic development. The stock markets are often unstable and can serve as an asset for the investors if become predictable. It will also be significantly beneficial for the country because the economy of the country will grow with better investment decisions, leading to a stable economy. The development of a financial model for the prediction of stock prices with high accuracy is very important for any stock market investment. There are many techniques available at present for forecasting the prices of the securities, yet statistical techniques are considered the most powerful and accurate among all. Stock markets are often very volatile to predict accurate stock prices. The volatile nature of the stock markets enforces the investors towards the diversification of risk to a trade-off between risk and return.

With a population of about 221.33 million, Pakistan stands at 5th rank among the most populated countries in the world. As compared to the other Asian counties, very few studies have been conducted using different machine learning statistics and deep learning techniques for forecasting and portfolio optimization. From a financial and economic point of view, Pakistan is struggling to adopt new techniques and tools for forecasting and analysis (Usmani, *et al.*, 2016). A very few studies have been conducted in Pakistan to forecast the stock prices using advanced statistical machine learning methods and Usmani (ibid) argues the use of various statistical machine learning techniques for predicting stock markets.

The study addresses the problem emphasized by Usmani (ibid) through an effort to forecast the KSE100 index using the most extensively applied financial time-series predicting methods. Pakistan's stock market is highly volatile and requires an optimized approach to diversify the risk by constructing an optimized portfolio. There are several ways to construct a portfolio that provides a trade-off between risk and return. The study uses Autoregressive Integrated Moving Average (ARIMA), Global Minimum Variance (GMV), and ARIMA – GMV hybrid approaches to construct a portfolio. ARIMA is used to forecast whereas the GMV technique offers amounts of weights to be allocated to all the securities alongside the covariance within a portfolio to get minimum value at risk using past values whereas ARIMA – GMV approach used the past, present, and future values to assign weights to construct optimize the portfolio.

The study significantly contributes by assisting the investors in the application of the more reliable and accurate forecasting approaches being used in advance statistics through machine learning approaches against the traditionally used methods. The study also motivates the investors to apply the advanced financial modeling approaches to optimize the portfolio for analysis and decision-making processes. The study has been able to empirically demonstrate the capacity of the ARIMA – GMV model to develop an optimum portfolio having capable of minimizing the overall risk of the portfolio to protect from the worst-case-losses. The application of advanced machine learning statistical techniques is also a small contribution towards the deficit in the advance statistical learning research in Pakistan as compared to its neighboring countries.

The research question imposed in this study investigates "Does ARIMA – GMV approach outperform the GMV technique in portfolio optimization"

To focus on the research problem, the study tries to achieve the following research objectives:

- Forecast the Portfolio Prices using ARIMA
- Construct a portfolio based on the GMV approach on the past stock prices
- Construct a portfolio based on the ARIMA – GVM approach on the past and forecasted stock prices
- Compare both the portfolio performance for evaluation of a better financial model for Portfolio Optimization

The remaining part of the document is arranged in the following way:

- Section 2 highlights the relevant work conducted in the past. This section provides the basis for forecasting and portfolio optimization using ARIMA and GMV approaches.
- Section 3 describes the methodology used in this study to run the experiment and address the research question by achieving the set research objectives.
- Section 4 presents the design specification of the models used in the study for forecasting and calculating individual stock weights using ARIMA and GMV approaches.
- Section 5 shows the implementation and evaluation of the study. The results are presented and discussed in detail in this section.
- Section 6 describes the key findings of the study, limitations, and future work of the study.

2. Related Work

The prediction of the stock prices and the optimization of the portfolios have been significantly studied in the academic literature especially after the introduction of the machine learning techniques in statistics (Ariyo, Adewumi and Ayo, 2014). Several studies have been conducted over the past in developed and developing countries for the application of the different statistical machine learning, and other techniques. Some of the studies are discussed below:

2.1. Auto-Regressive Integrated Moving Average (ARIMA)

Many studies have been conducted on the Asian Stock exchanges using different forecasting techniques, for example, Hasan, Majumder, and Hossain (2019) studied to develop a most accurate model for forecasting the stock prices using Feed Forward Neural Networks (FFNN), Linear model, Holt-Winters model, Holt-Winters model, and Autoregressive Integrated Moving Average (ARIMA). There was a total of 35 stocks listed at the Dhaka Stock Exchange. Results obtained from the study indicated that the ARIMA performed better than the remaining forecasting models on average. Results obtained for the Stock Index forecasting indicated that the FFNN performed better than the other models but determining the appropriate features of a neural network is very difficult and it is also very time consuming to determine the proper model when data and features increases (Kamruzzaman, Khudri and Rahman, 2017). Other previous studies (Kamruzzaman, et al., 2017; Bose, Uddin and Islam, 2014) performed on the Dhaka Stock Exchange for stock forecasts have also produced accurate results using ARIMA models.

Sharaff and Choudhary (2018) also studied prediction models for stock forecasting on the Indian stock exchange by using stochastic models such as Holt-Winters, Artificial Neural Network, RNN, and ARIMA models. The study was performed on the closing prices of the stocks listed at the Standard & Poor and Bombay Stock Exchange Sensex (S&P BSE). The models were evaluated using mean absolute error. The results indicated that ANN produced better prediction results than other models followed by ARIMA, but the study was not able to indicate the lag selection criteria and considered lagged values based on the ARIMA lagged values. A similar study was also conducted by Yadav and Sharma (2018) using forecasting methods, ETS, Naïve, Snaive, BoxCox, Meanf, Neural Network, and ARIMA for better stock market prediction models. Stock prices were taken from the last seventeen years from S&P BSE. The study concluded that the ETS outperformed all the other models in terms of Mean error followed by NN whereas a critical review of the study indicates that the NN and ARIMA have performed better in terms of other model evaluation approaches such as RMSE, MAE, MASE, and ACFI.

Sable, Shivani, and Goel (2019) conducted a study on the different forecasting techniques over the years including studies conducted on the Pakistani, Indian, and Bangladesh Stock Exchanges. The purpose of the study was to highlight the major traditional, machine learning, and deep learning techniques around the world because stock price prediction is one of the crucial and challenging tasks. The study proposed different methods for forecasting including the most widely used forecasting technique ARIMA because it has better statistical properties than others. The study also indicated the application of the hybrid ARIMA Neural Network model to deal with the nonlinearity of the time-series. The study also indicated mean square error (MSE) is the extensively used model evaluation technique followed by mean accuracy and mean absolute error.

Du (2018) studied the BP neural network, ARIMA, and ARIMA-BP neural network hybrid models on the Shanghai Securities Composite Index. The data was collected for one year from January 3, 2017, to December 22, 2017. Out of 240 observations, 200 observations were used as training data and the remaining 40 as test data to avoid the chance of single sample prediction. The study provided evidence that the ARIMA technique is more flexible in time-series analysis. ARIMA uses the time-series and regression techniques together and therefore is widely used for forecasting models. Although, the results obtained from the hybrid ARIMA-BP neural network model are better than the ARIMA model and BP model the study only accounts for the effect of one year and also didn't describe the financial or political environments effecting stock prices which can lead towards unpredictability and uncertainty.

2.2. Global Minimum Variance (GMV)

In the past, many studies have been conducted on the optimization of a portfolio using a different machine, deep learning and statistical techniques have been applied under various environments but statistical techniques have been considered the most reliable and accurate measure for the portfolio optimization especially with the help of the advance technology in the statistics, the power of statistical measures have dramatically increased. Among the different portfolio techniques, the global minimum variance approach is based on the work of the Markowitz has been of great importance to the investment practitioners and researchers (Paskalis, 2019).

Bodnar, Parolya and Schmid (2018) studied the Global Minimum Variance portfolio in the high dimensional case from the random matric theory outcomes. A dramatic improvement in the results was noted, turning into a robust model (Bodnar, *et al.*, 2018). In another study, Yang, Couillet and McKay (2015) were able to develop the estimate of the realized portfolio risk that was reduced through online optimizing the decreased intensity. For both synthetic and real market data, the portfolio technique of optimization was shown via replications to beat current techniques (Yang, *et al.*, 2015). There is also evidence from many studies that inferences for the parameters fail arising in the portfolio optimization, but these phenomena only happen when there are high dimensional assets universes (Taniguchi, *et al.*, 2018).

The standard portfolio theory indicates that only those portfolios are efficient, constructed using a tangency portfolio but many empirical studies provide sufficient evidence that the portfolios created through the GMV technique were able to produce improved results as compared to the tangency portfolio, recommending to invest in the portfolio following GVM technique. That study had made contributions by estimating the weight allocations and returning the portfolio parameter (Kempf & Memmel, 2006).

In the GVM approach, the Uncertainty is a key element, having a negative impact on developing a portfolio though the covariance matrix. To deal with the uncertainty effect, Benoit, Bertrand, and Sessi established a rule while using the GVM approach while constructing a portfolio that the investors should use alternative estimators. There is lower turnover and variance in the robust portfolios and high sharp ratios against the minimum variance portfolios (Maillet, Hossain and Hasan, 2015) whereas Stepan, Yerima, and Teras were able to determine the estimated weights of the portfolio with an assumption, the log-returns are distributed normally. To allow for information and noninformational priors in the GMV, the model reparametrized. For the remaining models, the weights are distributed in the portfolios by explicit form (Bodnar, *et al.*, 2017).

2.3.**ARIMA – Global Minimum Variance**

A study was conducted in Thailand by Thuakhonrak, Rattagan, and Phoomvuthisarn (2019) to propose a machine trading strategy using time series models such as ARIMA, Holt-Winters, and portfolio optimization strategies, introduced by Harry Markowitz, for the Stock Exchange of Thailand. The data was collected from the Stock Exchange of Thailand 50 index. The top five stocks were selected for the analysis based on the sharp ratio. The results showed that the portfolio outperformed the market returns. The ARIMA provided better results than the HW model. The portfolio was constructed based on the historical values and did not account for the effect of the values derived from the ARIMA model. The study lacks in establishing a relationship between the ARIMA and Markowitz approach because the stocks were selected based on the sharp ratio and the portfolio is constructed using the Markowitz approach without integrating the effect of ARIMA while developing a financial model for machine learning trading. The financial modeling based on the ARIMA or Holt-Winters model injected with a global minimum variance approach would provide different results.

The portfolio is constructed to optimize the trade-off between risk and return. Since the investment is made for an exchange of future excess returns, therefore, the investment decisions should be made based on the expected future returns, but future returns are based on the previous observations. To get the expected accurate future returns, several machine learning and statistical techniques are used but statistical techniques are considered the more reliable techniques than machine learning or deep learning techniques because of the statistical power. Among all the statistical methods, ARIMA is considered the most appropriate and used widely for forecasting future values. Especially in predicting stock prices. On the other hand, investors are also not willing to fully depend on future values and wanted to minimize the risk of loss. For portfolio optimization, the mean-variance and global minimum variance approaches are used but the global minimum variance approach provides better results than the mean-variance because, in the GMV approach, the portfolio is constructed based on the appropriate weights against the individual portfolio stocks to minimize the overall variance of the portfolio.

3. **Methodology**

The Cross-Industry Standard Practice Data Mining (CRISP-DM) technique is used for this study because

3.1.**Data Understanding and Preparation**

The stock prices selected for the study are based on the market capitalization because larger stocks have more effect on the stock index. The six largest stocks are selected for the portfolio optimization based on the proportion of the market share from the Karachi Stock Exchange Index as indicated in Table 1. The KSE100 index is a list of the top hundred companies in Pakistan across all sectors of the economy. The KSE100 is also used as a benchmark for the evaluation and analysis in Pakistan.

In the following step, the data is prepared by checking the missing values and plotting stock prices using graphs. Once the data is normalized using the log function, it is explored using autocorrection function, partial autocorrelation function, and Dickey-Fuller test to check the autocorrelation, partial autocorrelation, and stationery of the data. Dickey-Fuller test is run on the log values and differentiated log values to check stationery for each stock and KSE 100. The presence of the serial correlation is checked with the help of the Ljung-Box test. The null hypothesis indicates that there is no serial correlation in the data whereas the alternative hypothesis provides evidence of the presence of the serial correlation in the data. The data is transferred into time-series. The period covered is January 1, 2014, until December 31, 2019.

3.2.**Modeling**

The data were split into a training and test datasets with a ratio of 80/20. The training dataset represents 80% data whereas test data represents 20%. It is a fair and equal distribution of the data because the first four years' data, as part of the training data, is used to build the model whereas the remaining one year is for evaluation of the model performance.

Three models were developed to address the question. The first ARIMA model is developed for forecasting the weekly stock prices of all the stocks, and the KSE100 index using training data set through auto ARIMA function in r studio.

The second model is developed to construct a portfolio based on the GMV approach using the training dataset. The weights are calculated for each stock using actual training data variances and based on these weights, the portfolio is constructed.

The last model is developed to build a portfolio based on the actual past values and forecasted values of the ARIMA model. The weights of the individual stocks in the portfolio are calculated using the GMV approach.

3.3. **Measure for Goodness of Fit:**

For evaluation of the ARIMA model, the following measures are adopted:

Mean Percentage Error: Mean percentage error (MPE) is applied in many studies for model appraisal. It is a mean of the percentage differences of actual values and predicted values to the actual values.

$$MPE = \frac{\sum_{i=1}^{n}(\frac{y_i - \hat{y}_i}{y_i})}{n} \times 100 \tag{1}$$

Mean Absolute Percentage Error: Mean Absolute Percentage Error (MAPE) is another measure of evaluation for model performance and is applied to get the average absolute percentage difference of actual values y_i from the predicted values.

$$MAPE = \frac{\sum_{i=1}^{n}|\frac{y_i - \hat{y}_i}{y_i}|}{n} \times 100 \tag{2}$$

A one-million-dollar portfolio is constructed based on the weights calculated based on the second and the third model. A comparison is made between the performance of both the portfolios and the change in weights due to the change of rate.

Research Methodology

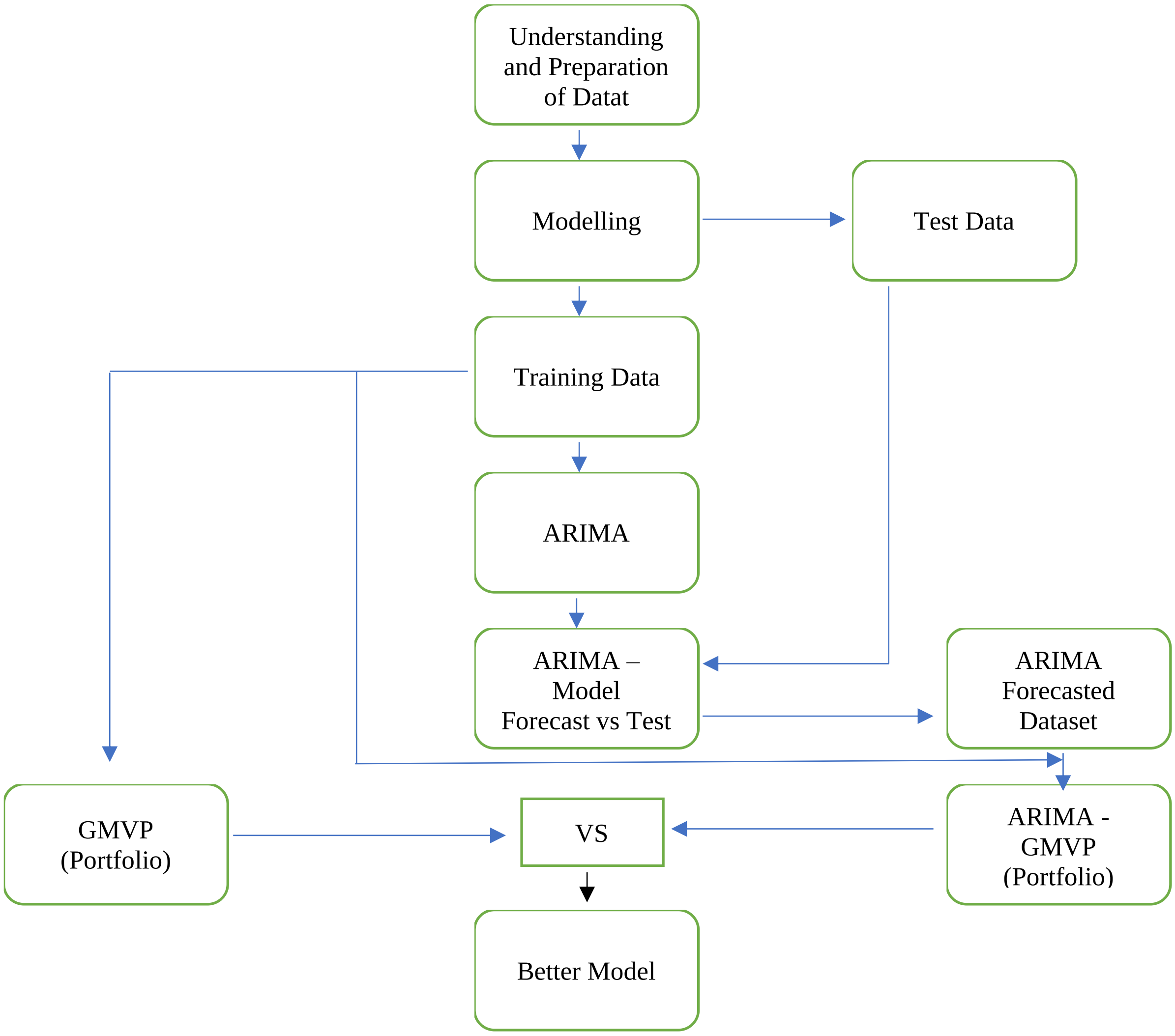

Figure 1: Research Methodology

4. **Design Specification**

4.1.**ARIMA Model:**

In the 1960s, Box and Jenkins developed a model for forecast values which is still most widely used for the statistical method for time-series forecast. The model was given name as an Autoregressive Integrated Moving Average (Wang and Niu, 2009). Autoregressive Integrated Moving Average (ARIMA) is an integrated model of the ARMA which is a combination of the Auto-Regressive (AR) and Moving Average (MA). Auto-Regressive indicates the linear function of expected value at any given time to its past observations. It can be expressed in the following equation form:

$$x_t = \delta + \emptyset x_{t-1} + \epsilon_t \tag{3}$$

Similarly, the Moving Average (MA) indicates that at any given point of time, the expected value is the linear function of the past error terms. The MA model can also be written in the following equation form

$$x_t = \mu + \theta \epsilon_{t-1} + \epsilon_t \tag{4}$$

Where δ (delta) and μ (mu) are constants

$\emptyset\ (phi)\ \theta\ (theta)\ are\ coefficients$

$\epsilon\ (epsilon)\ is\ error\ term$

The Auto-Regressive model (AR) and Moving Average model (MA) model also have the p and q parameters. The values of the parameters p and q of ARMA are calculated with the help of the autocorrelation and partial correlation functions by following the minimum criterion of AIC and BIC (Du, 2018).

The ARIMA model without any seasonality is represented by ARIMA (p,d,q),
Where p, d, and q are the number of autocorrelations, differences, and lag forecast errors.

The non-seasonal ARIMA model with parameters p, d, and q can be written in the following equation form:

$$x_t = \delta + \emptyset x_{t-1} + \epsilon_t - \theta \epsilon_{t-1} - \epsilon_t \tag{5}$$

$$x_t = \delta + \emptyset x_{t-1} - \theta \epsilon_{t-1} \tag{6}$$

Box and Jenkins (1970) also introduced the season model of the ARIMA which is an improved version of the non-seasonal ARIMA to deal with the effect of the seasonality in the data even after differencing the series and still, seasonality exists in the data (Wang and Niu, 2009). Seasonal ARIMA is represented by ARIMA (p,d,q)(P, D, Q)s, The parameters P, D, and Q are the seasonal orders of the model as parallel to its non-seasonal parameters (p,d,q). The parameters P, D, and Q represent the number of seasonal autocorrelation (SAR) term, seasonal difference, and seasonal moving average (SMA) term.

The seasonal model can be expressed in the form of the following equation;

$$[\nabla^{d}\nabla_{s}^{D}Y_{t} - \mu] = \frac{\theta(B)\vartheta(B^{s})}{\emptyset(B)\varphi(B^{s})} e_{t} \quad (7)$$

The study is using the Auto ARIMA function which is mostly used while performing statistical techniques using machine learning techniques. The Auto ARIMA function automatically detects the suitable parameters based on the data and forecast based on the best ARIMA model.

4.2. **Global Minimum Variance**

Global Minimum Variance (GMV) is a statistical technique applied for determining appropriate security weights in a portfolio to reduce the overall variance subject only to the constraint that the sum of the weights is equal to 1.
The GMV can be represented as

$$r_{na} = \alpha + \omega_{1}(r_{na} - r_{1}) + \omega_{2}(r_{na} - r_{2}) + \ldots. + \omega_{na-1}(r_{na} - r_{na-1}) + \epsilon_{t} \quad (8)$$
$$\omega_{na} = 1 - \sum_{i=1}^{na-1} \omega_{i} \quad (9)$$

Where α *is the mean of return*
r_{i} is the return of assets
ωi *is the weights of assets*
ϵ *is the error term*

4.3. **ARIMA – GMV**

ARIMA – GMV is a hybrid model of ARIMA and GVM approaches because in ARIMA – GMV the weights are allocated to the security based on the past observations and forecasted values derived from the ARIMA forecast.
The ARIMA – GMV can be represented as

$$r_{na} = \alpha + \omega_{\mathrm{n}}(r_{na+n} - r_{\mathrm{na+n}}) + \ldots + \omega_{0}(r_{na} - r_{\mathrm{na+0}}) + \ldots. + \omega_{na-1}(r_{na} - r_{na-1}) + \epsilon_{t} \quad (10)$$
$$\omega_{na+n} = 1 - \sum_{i=1}^{na+1} \omega_{i+n} \quad (11)$$

Where α *is the mean of return*
r_{na} is the return of assets including past, present, and future expected returns
ωi + n *is the weights of assets*
ϵ *is the error term*

5. **Implementation and Evaluation**

5.1.**Data Collection, Exploration, and Preparation**

The weekly closing prices of the KSE100 index and six stocks are collected over five years, 01 January 2015 to 31 December 2019. The overall representation of the sample is 20% of the KSE 100 index as indicated in Table 1.

Table 1: Stocks Selected for Analysis

S. No.	Stock	Symbol	Value	%Share
1	Muslim Commercial Bank	MCB	263,491,358.00	5.37%
2	Engro Foods	ENGRO	234,252,125.00	4.78%
3	Habib Bank Limited	HBL	152,681,658.00	3.11%
4	The Hub Power Company Limited	HUBC	138,811,450.00	2.83%
5	Oil & Gas Development Company Limited	OGDC	112,575,120.00	2.29%
6	Fauji Fertilizer Company Limited	FFC	81,328,205.00	1.66%
	Total		**983,139,916.00**	**20.04%**
	KSE		**4,905,558,354.69**	**100%**

With the help of exploratory analysis using graphs and statistical tests, the dataset contains no missing values but there is evidence that the data presented a trend as indicated in Figure 2. The stock prices of the ENGRO, FFC, HUBC, HBL, and KSE indicate the trend whereas OGDC and MCB show the trend and seasonality. There is a total of 260 weekly stock prices of each stock and KSE 100.

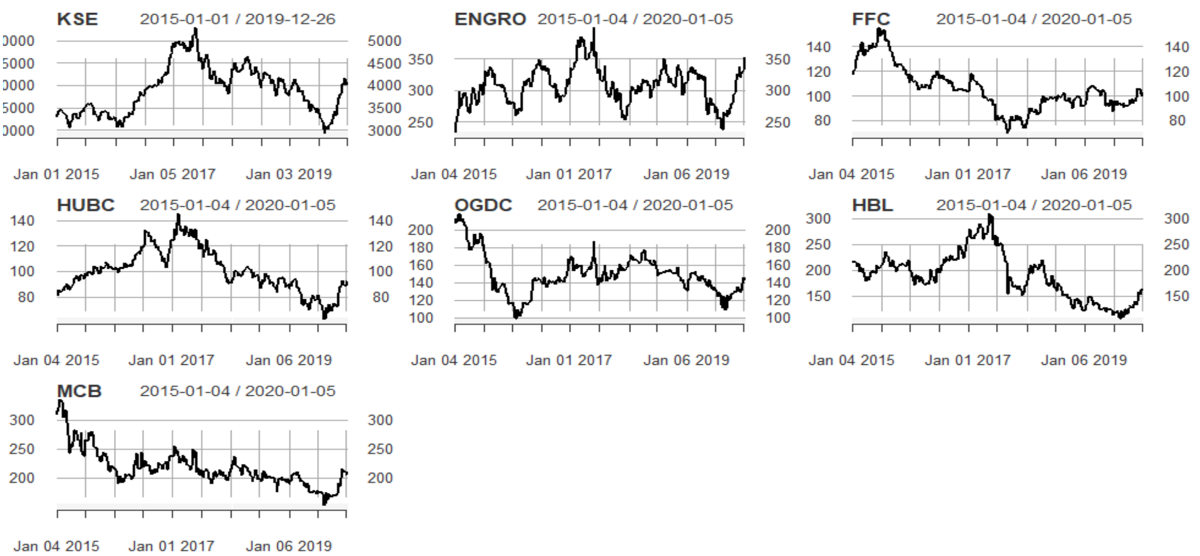

Figure 2: Weekly Closing Prices

The results obtained from the Ljung Box test as specified in Table 3 show that there is no serial correlation in the dataset. The results indicate that the observations are serially correlated in the first two stages. In the third stage, there is no serial correlation present in the data. So we can conclude that after taking the first difference of the log values, there is no serial correlation.

Table 3: Ljung Box Text Results (P-values)

	KSE	ENGRO	FFC	HUBC	OGDC	HBL	MCB
Value	$2.2x10^{-16}$	$2.2x10^{-16}$	$2.2x10^{-16}$	$2.2x10^{-16}$	$2.2x10^{-16}$	$2.2x10^{-16}$	$2.2x10^{-16}$
Log	$2.2x10^{-16}$	$2.2x10^{-16}$	$2.2x10^{-16}$	$2.2x10^{-16}$	$2.2x10^{-16}$	$2.2x10^{-16}$	$2.2x10^{-16}$
Diff. log	0.07	0.91	0.07	0.46	0.92	0.30	0.39

To normalize all the values, the data is transformed using log function on an initial 80% values which are used as training data. Tabel 4 gives Dickey-Fuller test results on the transformed data with and without the first difference. All the stocks show the non-stationarity in the data except MCB with a p-value of less than 5%. To convert the non-stationery data into stationary data, the first difference is taken. The Dickey-Fuller test is again run on the first differenced data. All the series in the data now represent the stationarity with a p-value of less than 5%.

Table 4: Dickey-Fuller Test Results

	KSE	**ENGRO**	**FFC**	**HUBC**	**OGDC**	**HBL**	**MCB**
W/o Diff.	0.9263	0.3026	0.6737	0.6782	0.2717	0.9432	0.0205
With Diff	0.0100	0.0100	0.0100	0.0100	0.0100	0.0100	0.0100

The Autocorrelation and Partial Autocorrelation functions are used to analyze the serial correlation between the values on the training data. Figure 3 and Figure 4 indicates the results of the autocorrelation and partial autocorrelation function. The autocorrelation function in Figure 3 and Figure 4 above significance level (Blue Dotted horizontal line above) indicates that the lag values are significantly correlated but the data at this point is still not stationary which means that we can’t determine the order of AR and MA processes using the ACF and PACF functions using Figure 3 and Figure 4. Using the Auto ARIMA function, the order of the ARIMA parameters can easily be calculated instead of using ACF and PACF plots to determine the manual order of parameters as indicated in Figure 5. Since the series indicated in figure 3 and Figure 4 are not white noise series so we cannot see the ACF function to see which parts spikes are the most significant. The values parameters of the ARIMA (p, q) and Seasonal ARIMA (p, q)(P, Q) are shown in Figure 5.

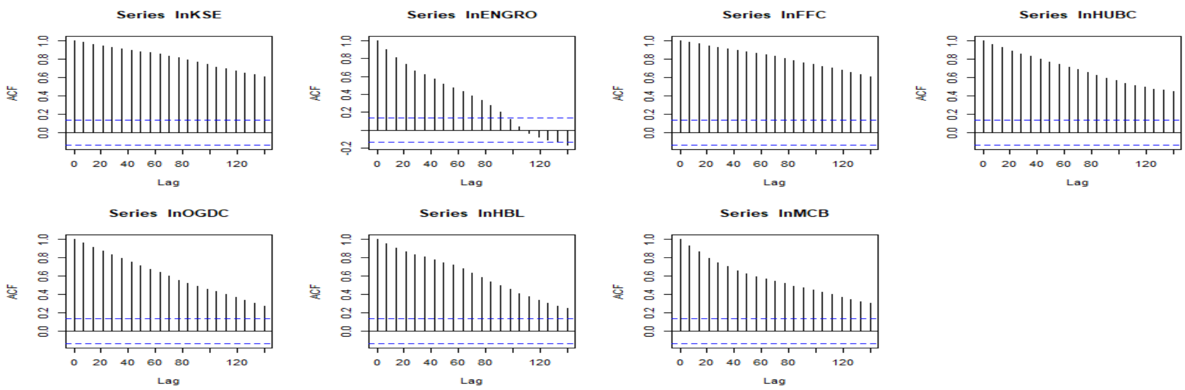

Figure 3: Autocorrelation Function

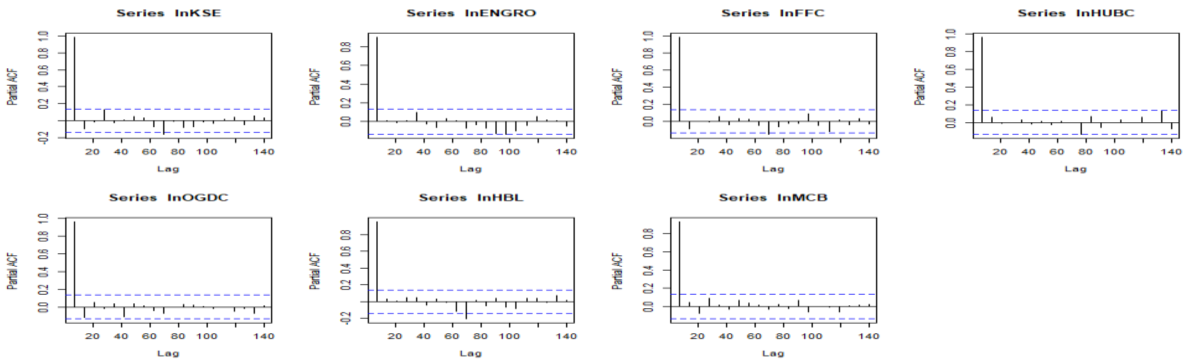

Figure 4: Partial Autocorrelation Function

5.2. **ARIMA Forests**

Using Auto ARIMA, the values of the KSE 100 index and other stocks included in the sample are forecasted for one year. The Results of the forecasts are shown in Figure 5 for the ARIMA models of KSE (0,1,3), ENGRO (1,1,1), FFC (0,1,0), HUBC (0,1,0), OGDC (3,1,2)(1,0,1)(52), HBL (0,1,0) and MCB (3,1,1)(1,0,0)(52). The models are shown in Figure 5 indicates the presence of the two seasonal models, one with simple seasonality whereas another seasonal model with drift. Each model in Figure 5 also indicates the confidence interval, highlighted in the blue shade.

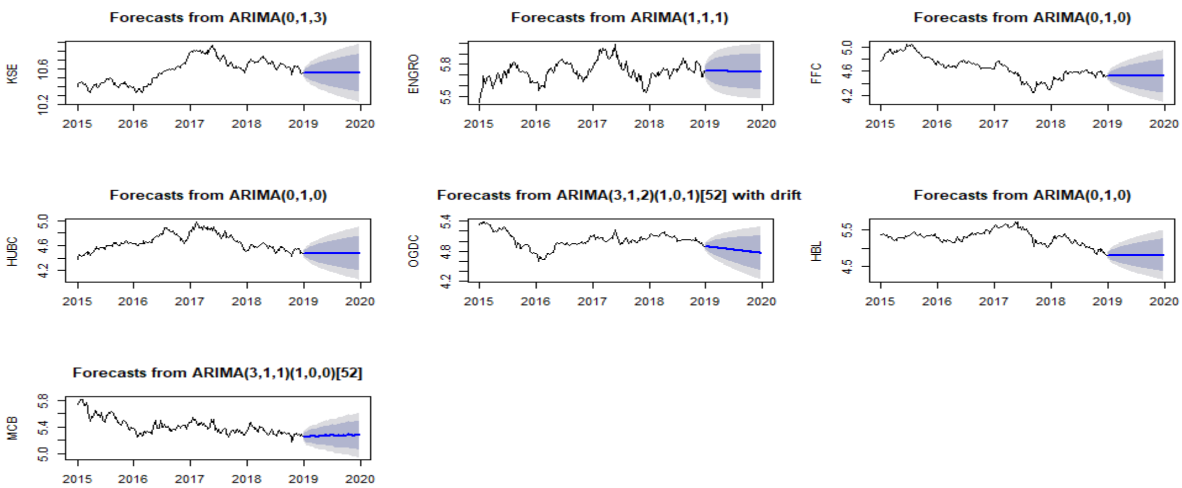

Figure 5: Weekly Closing Prices

Using the forecasted values of the training data, the mean percentile error of the model is measured by converting log values into nominal values using the exponential command in r studio. The results of the mean percentile are shown in Table 5. From the results, the forecasted values are on an average far below 10% which shows that model is performing well in general. Similar results have been obtained when the mean absolute percentage error method is used for model performance, the average overall results are below 10% which also indicates the goodness of the model for forecasting stock prices of KSE as supported by several studies (Tiwari, Bharadwaj and Gupta, 2017; NaitMalek, *et al.*, 2019).

Table 5: Mean Percentage Error and Mean Absolute Percentage Error

	KSE	**ENGRO**	**FFC**	**HUBC**	**OGDC**	**HBL**	**MCB**
MPE	-7.66	-5.50	6.41	-12.25	6.45	4.66	-5.40
MAPE	10.07	10.28	6.68	13.74	8.18	7.59	8.94

5.3. Global Minimum Variance

Using the regression technique of the GMV approach, the weights of the individual stocks in the portfolio are calculated to minimize the overall risk. The results of the GMV are displayed in Table 6. HBL, ENGRO, and OGDC are riskier with beta values of 0.5719, 0.5352, and 0.4536 whereas FFC is the least risker with a beta value of 0.2085 and therefore FFC is allocated more highest weightage 37.94% and HBL 0.23%. The overall portfolio beta is 0.3243 which is less than the average beta values of individual stocks 0.4270.

Table 6: Weights and Beta Values of Individual Stocks (GMV)

	GMP-Weights	**Stock Beta**
ENGRO	5.99%	0.5352
FFC	37.94%	0.2085
HUBC	37.74%	0.3439
OGDC	15.23%	0.4536
HBL	0.23%	0.5719
MCB	2.86%	0.4487
	100.00%	**0.3243**

5.4. ARIMA – GMV

Using the forecasted values of the individual stocks, the weights of the individual stocks in the portfolio are calculated and shown in Table 7. The weights of the stocks have changed as the beta values of the stock have changed except for HUBC. The weights allocated to the highest stock FFC have changed from 37.94% to 37.96 and for HBL from 15.23% to 15.24%.

Table 7: Weights and Beta Values of Individual Stocks (ARIMA-GMV)

	ARIMA-GMP Weights	Stock Beta
ENGRO	5.98%	0.5358
FFC	37.96%	0.2081
HUBC	37.74%	0.3441
OGDC	15.22%	0.4540
HBL	0.24%	0.5711
MCB	2.85%	0.4477
	100.00%0	**0.3242**

Based on the actual stock prices at the weekend price of stocks on December 31, 2019, the Change of Rate (COR) is measured against the previous year's weekly ending stock prices as on December 31, 2018, as highlighted in Figure 7. The change in individual weights (CIW) indicates relative change against the actual rate of Change (ROC) in each stock. The overall CIW is directly proportionate to the actual COR.

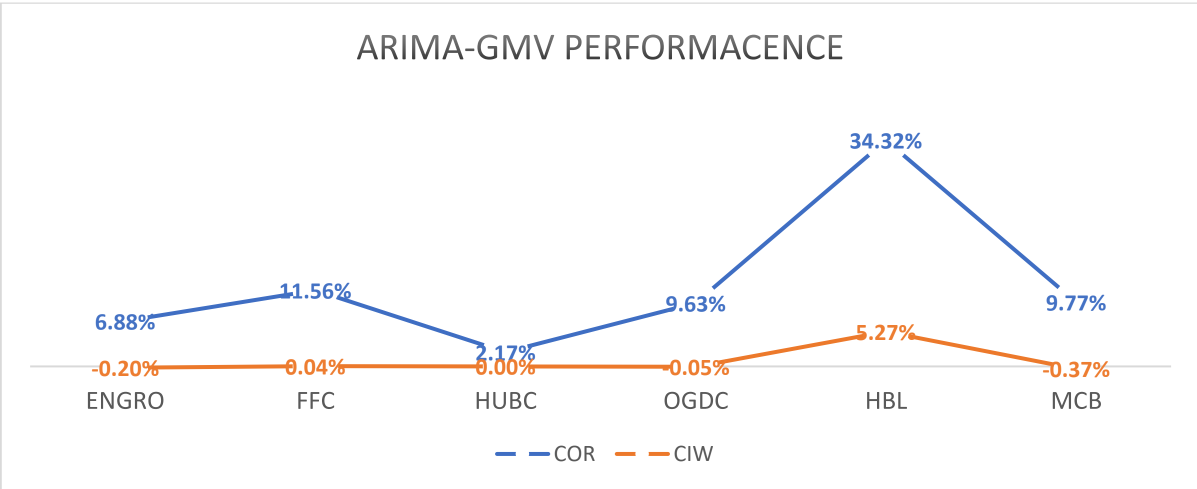

Figure 7: Change of Rate (COR) vs Change in Weights (CIW)

Based on an investment of $1,000,000, the actual stock performance of the portfolio is also measured in Table 8. In case I, the investment is made based on the GMV weights as on December 31, 2018, whereas in case II, the investment is made based on the ARIMA-GMV weights as on December 31, 2018. The overall portfolio investment is $1,074,467 in case I whereas $ 1,074,486 in Case II.

Table 8: Portfolio Performance (GMV vs ARIMA-GMV)

Case	ENGRO	FFC	HUBC	OGDC	HBL	MCB	Total
GMV – I	64,068	423,281	385,590	166,982	3,111	31,435	1,074,467
ARIMA-GMV – II	63,941	423,449	385,608	166,893	3,275	31,319	1,074,486

5.5. **Discussion**

Portfolio optimization is a way to minimize the risk of the overall investment by diversifying the risk among different investment options. The global minimum variance approach is one of the most widely used techniques to optimize the portfolio risk and return of portfolio investment. In the global minimum portfolio approach, the portfolio is constructed based on the variance of the past returns, and weights are allocated based on the relationship of the individual securities' returns' mean and covariance. But returns are expected to be received in the future and therefore the portfolio should not only be built on the past or future values but rather it should consider past, present, and expected future values.

Among the different forecasting techniques, ARIMA is the most widely and powerful statistical machine learning logarithm to predict stock prices. Several assumptions need to fulfill before ARIMA can forecast but the exploratory analysis shows that the initially retrieved data does not fulfill the assumption and therefore certain adjustments have been made in the data to use it for the ARIMA forecast. The data is transformed into the log values to normalize it and the first difference is taken to convert into stationary form and remove any serial correlation. Autocorrelation and Partial Autocorrelation functions are also plotted to check the serial correlation and the ARIMA parameters.

Auto ARIMA function is used to forecast the stock price because there are limitations with the use of the various approaches in calculating the ARIMA models which include the calculation of Akaike Information Criterion and Bayesian Information Criterion, autocorrelation and partial autocorrelation function plots. To overcome these limitations, the study used the auto ARIMA function in R forecasting packages (Nguyen, *et al.*, 2019). The Auto-ARIMA function provided the best forecasting ARIMA models for each stock. The Auto-ARIMA also forecasted the best models for the seasonality in the stock price.

The performance of each model is evaluated using the Mean Percentage Error and Absolute Mean Percentage Error. The results obtained from the MPE shows that all the models performed well with a mean accuracy error of less than 10% (NaitMalek, *et al.*, 2019; Tiwari, *et al.*, 2017) except HUBC with a 12.25% deviation from the actual values. Similar results are found in MAPE. Based on the model performance, the forecasted values of the stocks can be used for investment purposes.

The weights of the securities in constructing a portfolio are calculated based on the GMV approach only. The data used is the training data. The results clearly state that the stock with high beta is given fewer weights and vice versa. The overall beta value of the stock is less than the average values of the stocks. Weights are calculated based on the training data and a portfolio is constructed using the weights allocated to training data as indicated in Table 6 and Table 7 for comparison and analysis purpose with the Portfolio constructed based on the hybrid ARIMA-GMV approach. The weights allocated to the portfolio are slightly different than the weights calculated in the simple GMV approach. The individual weights in the ARIMA-GMV approach are directly proportionate to the performance of the ARIMA model performance.

The performance of the ARIMA-GMV approach as indicated in Figure 7 provides evidence that the weights of the individual stocks in the portfolio change as per expected values. The accurate the ARIMA model, the accurate weights will be allocated to the individual stocks in the portfolio. The change in the prices of the stocks as highlighted in Figure indicates that weights also change proportionately. The overall change in the weights is corresponding to the overall change in the stock prices. Table 7 shows that the performance of the ARIMA – GVM portfolio is better than the GVM portfolio.

ARIMA – GMV approach provides better results in financial modeling for portfolio optimization because ARIMA – GMV approach considers the accuracy of the ARIMA and GMV models. A combination of both the techniques provides a better cushion to the investors in protecting them from worst-case losses because the ARIMA model provides the basis for measurement of expected risk in the future and GVM diversify the overall risk of the individual securities by assigning appropriate weights to the individual securities to minimize the overall risk of the portfolio.

6. Conclusion and Future Work

Portfolio optimization has always been an interesting topic in academics and non-academic circles because stock markets are often unpredictable and due to the volatility and unpredicted nature of the stock markets, it is often difficult to forecast and get the desired results. An element of risk is always present while investing in securities because many securities are normally riskier than the others. The difference between the level of risk, influences the investors to go for such opportunities that provide better portfolios to protect them against worst-case-losses. The GMV method is used to decide the best weights of each stock for an optimized portfolio but the GMV method is based on past observations and does not consider the future expected observations. The study addresses the questions of whether the ARIMA-GMV approach provides a better financial model for portfolio optimization by forecasting the expected future values using the ARIMA approach and constructing an ARIMA – GMV portfolio based on the past actual and future expected values.

The results obtained from the study indicate that the ARIMA provides accurate results for the overall portfolio because the mean percentage error for the individual stock forecasted values is centrally tendency between -5.40% to 4.60% with an overall mean percentage error for the portfolio is less than 1%. Based on the level of overall accuracy, the forecasted values are used for the calculation of appropriate weights for portfolio optimization using the ARIMA – GMV approach. The case of a $ 1 million portfolio has backed the argument that the ARIMA – GMV portfolio performs better than the traditional GMV portfolio. Based on the evidence, the study proposes to use the ARIMA - GMV approach to determine the optimum weights for portfolio optimization.

The study has the following limitations:

- Very limited literature is available related to the use of machine learning statistical techniques in Pakistan and the use of the hybrid ARIMA – GMV approach in financial modeling for portfolio optimization.

- ARIMA – GMV approach mainly depends on the ARIMA model for forecasted values. The inaccuracy of the ARIMA model can lead the results towards poor portfolio optimization but still, the presence of past data will not affect the overall results.

- Political instability within the country has made the stock market more volatile over the past two years which makes it difficult to forecast accurate results.

The following future studies are suggested:

- Use of ARIMA – GMV to optimize the cross-border securities portfolio to absorb the risk of market volatility due to political instability and diversify the risk among different international markets.

- The use of different non – linear approaches to construct a hybrid GMV portfolio because markets are generally non – learn and different non – linear machine learning and deep learning techniques can also assist in forecasting non – linear patterns for portfolio optimization.

References

Ariyo, A. A., Adewumi, A. O. and Ayo, C. K. (2014) ‘Stock Price Prediction Using the ARIMA Model’, in *2014 UKSim-AMSS 16th International Conference on Computer Modelling and Simulation, Computer Modelling and Simulation (UKSim)*. Cambridge, United Kingdom, 26-28 March 2014, pp. 106–112, IEEE Xplore. doi: 10.1109/UKSim.2014.67.

Banerjee, D. (2014) ‘Forecasting of Indian stock market using time-series ARIMA model’, in *2014 2nd International Conference on Business and Information Management (ICBIM)*. Durgapur, India, 09-11 January 2014, pp. 131–135, IEEE Xplore. doi: 10.1109/ICBIM.2014.6970973.

Bodnar, T., Mazur, S. and Okhrin, Y. (2017) ‘Bayesian estimation of the global minimum variance portfolio’, *European Journal of Operational Research*, 256(1), pp. 292–307. doi: 10.1016/j.ejor.2016.05.044.

Bodnar, T., Parolya, N. and Schmid, W. (2018) ‘Estimation of the global minimum variance portfolio in high dimensions’, *European Journal of Operational Research*, 266(1), pp. 371–390. doi: 10.1016/j.ejor.2017.09.028.

Bodnar, T. and Schmid, W. (2008) ‘A test for the weights of the global minimum variance portfolio in an elliptical model’, *Metrika*, 67(2), pp. 127-143. doi: 10.1007/s00184-007-0126-7.

Bose, T. K., Uddin, M. R. and Islam, M. W. (2014) ‘Measuring and Comparing the Efficiency of Dhaka Stock Exchange and Chittagong Stock Exchange’, *International Journal of Scientific and Research Publications,* 4(3), pp. 1-14.

Du, Y. (2018) ‘Application and analysis of forecasting stock price index based on combination of ARIMA model and BP neural network’, in *2018 Chinese Control And Decision Conference (CCDC), Chinese Control And Decision Conference (CCDC),* Shenyang, China, 9-11 June 2018, pp. 2854–2857. doi: 10.1109/CCDC.2018.8407611.

Frahm, G. and Memmel, C. (2010) ‘Dominating estimators for minimum-variance portfolios’, *Journal of Econometrics*, 159, pp.289-302, Science Direct. doi: 10.1016/j.jeconom.2010.07.007.

Glombek, K. (2014) ‘Statistical Inference for High-Dimensional Global Minimum Variance Portfolios’, *Scandinavian Journal of Statistics,* 41(4), pp. 845-865.

Haider, A. S. (2009) ‘Forecasting Dhaka Stock Exchange (DSE) return: An Autoregressive Integrated Moving Average (ARIMA) approach’, *North South Business Review,* 3(4), pp. 36-54.

Kamruzzaman, M., Khudri, M. M. and Rahman, M. M. (2017) ‘Modeling and Predicting Stock Market Returns: A Case Study on Dhaka Stock Exchange of Bangladesh’, *Dhaka University Journal of Science,* 65(2), pp. 97-101.

Kempf, A. and Memmel, C. (2006) ‘Estimating the Global Minimum Variance Portfolio’, *Schmalenbach Business Review*, 58(4), pp. 332-348, Business Source Ultimate. doi: 10.1007/BF03396737.

Maillet, B., Tokpavi, S. and Vaucher, B. (2015) ‘Global Minimum variance portfolio optimisation under some model risk: A robust regression-based approach’, *European Journal of Operational Research,* 244(1), pp. 289-299. doi: 10.1016/j.ejor.2015.01.010.

Majumder, M. M. R., Hossain, M. I. and Hasan, M. K. (2019) ‘Indices prediction of Bangladeshi stock by using time series forecasting and performance analysis’, in *2019 International Conference on Electrical, Computer and Communication Engineering (ECCE)*. Cox's Bazar, Bangladesh, pp. 1-5. doi: 10.1109/ECACE.2019.8679480.

NaitMalek, Y., Najib, M., Bakhouya, M. and Essaaidi, M. (2019) ‘On the Use of Machine Learning for State-of-Charge Forecasting in Electric Vehicles’, in *2019 IEEE International Smart Cities Conference (ISC2), Smart Cities Conference (ISC2)*. Casablanca, Morocco, 14-17 October 2019, pp. 408–413, IEEE Xplore. doi: 10.1109/ISC246665.2019.9071705.

Nguyen, H. V., Naeem, M. A., Wichitaksorn, N. and Pears, R. (2019) ‘A smart system for short-term price prediction using time series models’, *Computers and Electrical Engineering*, 76, pp. 339–352, Science Direct. doi: 10.1016/j.compeleceng.2019.04.013.

Glabadanidis, P. (2019) ‘An Exact Test of the Improvement of the Minimum Variance Portfolio’, *International Review of Finance,* 19(1), pp. 45 - 82. doi: 10.1111/irfi.12173.

Tiwari, S., Bharadwaj, A. and Gupta, S. (2017) ‘Stock price prediction using data analytics’, in 2017 International Conference on Advances in Computing, Communication and Control

(ICAC3). Mumbai, India, 1-2 Dec. 2017, pp. 1–5, IEEE Xplore. doi: 10.1109/ICAC3.2017.8318783.

Sable, R., Goel, S. and Chatterjee, P. (2019) 'Empirical Study on Stock Market Prediction Using Machine Learning', in *2019 International Conference on Advances in Computing, Communication and Control (ICAC3)*. Mumbai, India, 20-21 Dec. 2019, pp. 1-5, IEEE Xplore. doi: 10.1109/ICAC347590.2019.9036786.

Sharaff, A. and Choudhary, M. (2018) 'Comparative Analysis of Various Stock Prediction Techniques', in 2018 2nd International Conference on Trends in Electronics and Informatics (ICOEI). Tirunelveli, India, 11-12 May 2018, pp. 735–738, IEEE Xplore. doi: 10.1109/ICOEI.2018.8553825.

Taniguchi, M., Shiraishi, H., Hirukawa, J., Solvang, H. K. and Yamashita, T. (2018) Statistical portfolio estimation, 3rd edn. London: CRC Press, Taylor & Francis Group.

Thuankhonrak, P., Rattagan, E. and Phoomvuthisarn, S. (2019) 'Machine Trading by Time Series Models and Portfolio Optimization', in 2019 4th International Conference on Information Technology (InCIT), Information Technology (InCIT). Bangkok, Thailand, 24-25 Oct. 2019, pp. 217–222, IEEE Xplore. doi: 10.1109/INCIT.2019.8912015.

Usmani, M., Adil, S. H., Raza, K. & Ali, S. S. A. (2016) 'Stock market prediction using machine learning techniques', in 2016 3rd International Conference on Computer & Information Sciences (ICCOINS). Kuala Lumpur, Malaysia, 15-17 Aug. 2016, pp. 322-327, IEEE Xplore. doi: 10.1109/ICCOINS.2016.7783235.

Wang, W. and Lv, Y. (2013) 'A study of the USDX based on ARIMA model — A correlation analysis between the USDX and the Shanghai index', in 2013 3rd International Conference on Consumer Electronics, Communications & Networks. Xianning, China, 20-22 Nov. 2013, pp. 49 – 53, IEEE Xplore. doi: 10.1109/CECNet.2013.6703269.

Wang, W. and Niu, Z. (2009) 'Time Series Analysis of NASDAQ Composite Based on Seasonal ARIMA Model', in 2009 International Conference on Management and Service Science. Wuhan, China, 20-22 Sept. 2009, pp. 1–4, IEEE Xplore. doi: 10.1109/ICMSS.2009.5300866.

Wied, D., Ziggel, D. and Berens, T. (2013) 'On the application of new tests for structural changes on global minimum-variance portfolios', Statistical Papers, 54(4), pp. 955 – 975. doi: 10.1007/s00362-013-0511-4.

Yadav, S. and Sharma, K. P. (2018) 'Statistical Analysis and Forecasting Models for Stock Market', in 2018 First International Conference on Secure Cyber Computing and Communication (ICSCCC), Jalandhar, India, 15-17 Dec. 2018, pp. 117–121, IEEE Xplore. doi: 10.1109/ICSCCC.2018.8703324.

Yang, L., Couillet, R. and McKay, M. R. (2015) 'A Robust Statistics Approach to Minimum Variance Portfolio Optimization', IEEE Transactions on Signal Processing, 63(24), pp. 6684–6697, IEEE Xplore. doi: 10.1109/TSP.2015.2474298.