

https://dspace.org/

Forecasting volatility in the South African stock market: A comparison of methods

Harrilall, Ushir

2014-08-06
http://hdl.handle.net/10539/15124

Downloaded from a DSpace repository. https://dspace.org/

Forecasting Volatility in the South African Stock Market: A Comparison of Methods

Ushir Harrilall
376656
Supervisor: Mr. Y. Seetharam

A research report submitted to the School of Economic and Business Sciences, Faculty of Commerce, Law and Management, University of the Witwatersrand, in partial fulfillment (50%) of the requirements for degree of Master of Commerce in Finance.

Johannesburg, South Africa

February 2014

DECLARATION

I, Ushir Harrilall, declare that this research report is my own unaided work. It is submitted in partial fulfilment of the requirements for the degree of Master of Commerce in Finance at the University of the Witwatersrand, Johannesburg. It has not been submitted before for any degree or examination at this or any other university.

ACKNOWLEDGEMENTS

I would like to thank the following people for their unrelenting support and encouragement:

My family, for the sacrifices they have made as well as for reminding me of what's important;

Mr.Y. Seetharam, my supervisor, for the ongoing advice and for introducing me to neural networks;

Justin Colyn, a fellow Master's student

And lastly, the entire Advanced Analytics team at Deloitte for their input and advice surrounding artificial intelligence forecasting techniques.

Definitions of Terms and Abbreviations

SAVI – South African Volatility Index

TDNN – Time-Delay Neural Network

EMH – Efficient Market Hypothesis

TDNN – Time-Delay Neural Network

CAPM – Capital Asset Pricing Model

ARMA- Autoregressive Moving Average

ARIMA – Autoregressive Integrated Moving Average

SES- Simple Exponential Smoothing

EWMA- Exponentially Weighted Moving Average

HIS- Historical Volatility Models

ARCH- Autoregressive Conditional Heteroskedasticity

GARCH – Generalised Autoregressive Conditional Heteroskedasticity

RMSE – Root Mean Squared Error

EBP – Error Backpropagation

Forecasting Volatility in the South African Stock Market: A Comparison of Methods

ABSTRACT

Volatility prediction has become a crucial task in the appraisal of assets and risk management. Increased financial regulation with tougher capital standards and additional capital buffers has made understanding volatility in financial markets even more imperative. This study investigates and compares various forecasting techniques in their ability to forecast the South African Volatility Index (SAVI). In particular, a time-delay neural network's forecasting ability is compared to those of more traditional methods. A comparison of the residual errors of all the forecasting tools used suggests that the time-delay neural network and the historical average model have superior forecasting ability over traditional forecasting models, with the naive historical average model having only slight superior forecasting ability than the neural network. From a practical perspective, this suggests that the historical average model is the best forecasting tool used in this study, as it is less computationally expensive to implement compared to the neural network. This result should however be interpreted with caution as only historical values of the SAVI were used as inputs to the neural network. In addition, the neural network may be better suited if the sample period were longer. Furthermore, the results suggest that the SAVI is extremely difficult to forecast, with the volatility index being purely a gauge of investor sentiment in the market, rather than being seen as a potential investment opportunity.

Table of Contents

1. Introduction 1
1.1 Background to the Topic 2
1.2 Research Problems and Objectives 5
1.2 Feasibility of Study 6
1.4 Hypothesis and Theoretical Framework 7
1.5 Summary of Findings 10
2 Literature Review 12
2.1 The Efficient Market Hypothesis and Forecasting 12
2.2 Forecasting Volatility Using Historical Data 16
2.3 Forecasting Volatility using Traditional Methods 17
2.4 Forecasting Volatility using Artificial Intelligence 22
2.5 The South African Volatility Index – An Introduction 23
2.5.1 The Benefits of the new SAVI 27
2.6 Summary 27
3. Research Methodology 28
3.1 Data 28
3.2 Historical Volatility Models 29
3.2.1 Historical Average Model (HAM) 30
3.2.2 Simple Exponential Smoothing (SES) 30
3.2.3 Exponentially Weighted Moving Average (EWMA) 31
3.3 Autoregressive and Heteroskedastic Models 31
3.3.1 Autoregressive Moving Average (ARMA) 31
3.3.2 Autoregressive Conditional Heteroskedasticity (ARCH) 32
3.3.3 Generalised Autoregressive Conditional Heteroskedasticity (GARCH) 32

3.4 Neural Network Model Construction 33
3.4.1 Classification versus Regression 33
3.4.2 Multilayer Perceptron (MLP) 34
3.4.3 Time Delay Neural Network (TDNN) 35
3.4.4 Learning Environment: Variable Selection 35
3.4.5 Data Pre-processing 37
3.4.6 Network Training, Testing and Validation Sets 38
3.4.7 Number of Hidden Layers 40
3.4.8 Number of Hidden Neurons 41
3.5. Experiment 41
3.5 Running the Multilayer Perceptron Programme 42
4. Empirical Results 42
4.1 Examining the Univariate Properties 43
4.2 Normality and Stable Moments 44
4.3 Stationarity and Unit Root tests 46
4.4 Determining the ARMA structure of the Data Generating Process 48
4.4.1 ARMA Equation Diagnostics 50
4.5 Testing for ARCH Effects 55
4.6 Historical Average Model (HAM) 57
4.7 Exponential Smoothing 57
4.8 Exponentially Weighted Moving Average 58
4.10 ARMA Model 59
4.11 ARCH Model 60
4.12 GARCH 61
4.13 TDNN 62
5. Conclusion 66
5.1 Summary of Results 66
5.2 Recommendations for Further Research 67
Reference List 69
Appendix A1 77
Appendix A2 78

Appendix B 80
Appendix C 83
Appendix D 85

List of Figures:

Figure 1: Weekly volatility of the SAVI from 22 July, 2007 to 22 July, 2013. 3
Figure 2: Average volatility of stock prices before and after announcement day 3
Figure 3: Typical architecture of a multilayer perceptron 8
Figure 4: Typical architecture of a time delay neural network 9
Figure 5: Flow diagram for creating and testing the model 10
Figure 6: FTSE/JSE Top40 index level and its volatility. 24
Figure 7: The SAVI and the new SAVI. The FTSE/JSE Top40 index level is also plotted. 25
Figure 8: Correlogram of the return series 43
Figure 9: Histogram of the return series 45
Figure 10: A plot of the daily return series 46
Figure 11: Graphical view of the inverse roots of the AR and MA characteristic polynomial 51
Figure 12: Graphical view of the actual and ARMA model correlogram 52
Figure 13: Impulse response function of ARMA model 53
Figure 14 : Graphical representation of TDNN 62
Figure 15: Algorithm and progress of TDNN 63
Figure 16: Best validation performance 63
Figure 17: Plot of the error autocorrelation function 64
Figure 18: Plot of inputs, targets and errors against time 64

List of Tables:

Table 1: Summary statistics for the return series....45
Table 2: Augmented Dickey Fuller test statistics....47
Table 3: Phillips-Perron test statistics....48
Table 4: AIC Values for various lag order combinations....49
Table 5:Estimation output - ARMA (4, 4) model....49
Table 6: Tabular view of the inverse roots of the AR and MA characteristic polynomial....51
Table 7: Tabular view of the actual and ARMA model correlogram....52
Table 8: Tabular view of ARMA impulse response function....54
Table 9: Correlogram of the residuals - ARMA model....55
Table 10: Breusch-Godfrey serial correlation LM test....56
Table 11: Single Exponential Smoothing results....58
Table 12: EWMA results....58
Table 13: ARMA results....59
Table 14: ARCH results....60
Table 15: GARCH (1, 1) results....61
Table 16: Summary of results....67

1. Introduction

Share prices are subject to large, unexpected inclines or declines. These large and unexpected movements could be linked to market crashes, which even though are baffling to understand, have had several theories developed to attempt to explain such phenomena (Antonakakis & Scharler, 2012). The volatility of financial markets usually exhibits a continuous and clustering character (Granger & Poon, 2005). In addition, the time-varying characteristics of stock markets which date back to the studies conducted by Mandelbrot (1963) and Fama (1965) is further motivation for volatility prediction, as well as well as further evidence of volatility clustering.[1]

Volatility prediction has become a crucial task in the appraisal of assets and risk management. Increased financial regulation with tougher capital standards and additional capital buffers have made understanding volatility in financial markets even more imperative. In addition, most derivative securities are affected by volatility, with most risk management models used by financial institutions and regulators relying on time-varying volatility as a key input (Brownlees, Engle & Kelly, 2012). Granger and Poon (2003, p.478) argue that when volatility is interpreted as 'uncertainty' in the market, it becomes a crucial input to investment decisions and portfolio formation, with investors and portfolio managers both having particular levels of risk which they can tolerate.

There is an important relation between the uncertainty of financial markets and public confidence. Policy makers often resort to market estimates of volatility as an indicator for the susceptibility of financial markets and the economy (Granger & Poon, 2003). Extreme volatility in financial markets which is often associated with financial crises has plagued many economies over the years and has been a catalyst for the growing interest in volatility forecasting tools.

Furthermore, the interconnected nature of financial markets, together with its complexity, has created the need for analytical tools that allow for a large number of market variables to be used

[1] The authors found that large (small) price changes tend to be followed by large (small) price changes, which is indicative of volatility clustering.

to explore interrelationships in financial markets; in particular, Artificial Neural Networks (hereafter referred to as a neural network). A neural network has the ability to learn nonlinear mappings between inputs and outputs. It thus may have the ability to predict stock market volatility. Neural networks can be trained to perform a variety of financial related tasks. This study analyses and examines the use of neural networks as well as traditional linear and non-linear methods as a forecasting tool. Specifically, a comparison is done between traditional forecasting methods and a Time-Delay Neural Network with regard to the various models' ability to forecast volatility.

1.1 Background to the Topic

Financial time series prediction is a complicated task which may be attributed to the following reasons:

1) Hellstrom and Holmstrom (1998, p.30) posit that financial time-series often behave like a total or 'near to random walk' process. Theoretically, this would make the prediction of a financial time series impossible (Hellstrom & Holmstrom, 1998).
2) Financial time-series are subject to regime shifting which implies that the statistical properties of the time series vary across different points in time (Hellstrom & Holmstrom, 1998).
3) Chiu, Lee and Lu (2009) state that financial time-series are very noisy, which implies that there are a lot of unpredictable variations.
4) In the long run, a new prediction technique influences the process to be predicted (Hellstrom & Holmstrom, 1998).

This study makes use of the SAVI as a means to examine volatility in the South African stock market, and is explored in detail later. Figure 1 below depicts the weekly[2] volatility of the SAVI from 2007 to 2013. The figure depicts that volatility changes greatly through time, consistent with the aforementioned points. If the market exhibited no volatility, throughout the period from

[2] Further empirical analysis throughout this study is done using daily data.

2007 to 2013, the graph in Figure1 would have been flat. There is evidence of the index values fluctuating greatly across certain weeks as well as periods where the index is calmer. Furthermore, Fair (2000) and Donders and Vorst (1996) claim that financial time-series are heavily impacted by specific events occurring on a certain day.

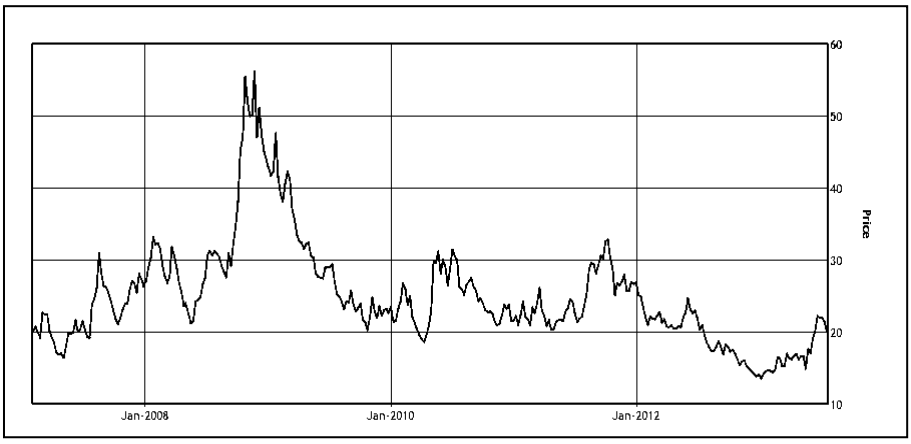


Figure 1: Weekly volatility of the SAVI from 22 July, 2007 to 22 July, 2013. (Source: McGregor BFA)

Figure 2 below shows a case whereby the volatility of stock prices increases as the announcement day of firm specific news approaches, and is then followed by a subsequent sharp decline in volatility after the announcement day.

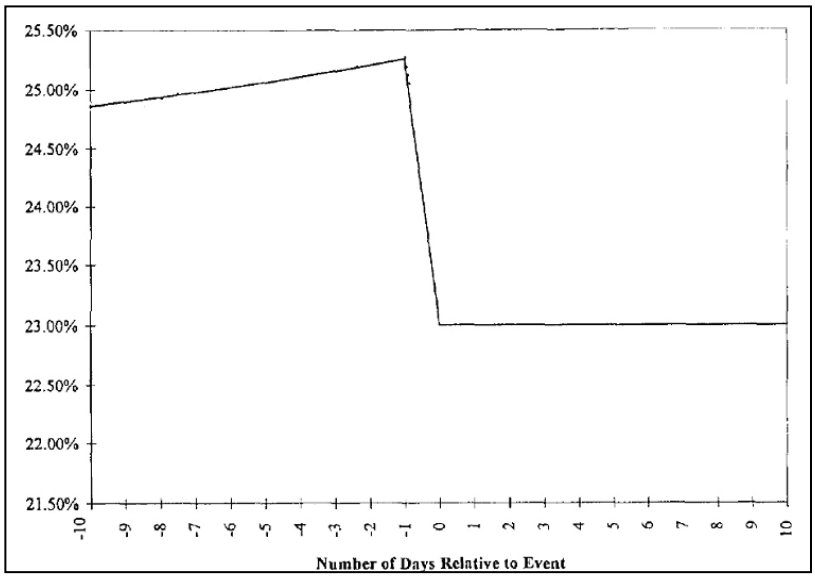


Figure 2: Average volatility of stock prices before and after announcement day (Donders & Vorst, 1996)

The ability to forecast the volatility of financial markets is essential to analysts. Dixit, Roy and Uppal (2013) argue that measuring market volatility became of paramount importance after the global stock market crash in 1987. Financial market volatility affects the entire economy with most definitions of a financial crisis being associated with a rare event, which is paralleled with a remarkable loss in market value. Crises tend to occur unexpectedly and events such as natural disasters, terrorist attacks and bad news regarding fundamentals may all be catalysts for such events. Financial reporting scandals which have plagued the United States' (US) financial markets, as well as the terrorist attacks on September 11, 2001, are events which caused turmoil not only in the US, but had ripple effects on several continents (Granger & Poon, 2003).

Assessing the values of financial indicators is complicated by complex interconnections, which are often not intuitive and sometimes convoluted, making volatility prediction an extremely difficult task. Authors have however, used various models in order to forecast volatility. These models can be split into two broad groups: stochastic models[3] and deterministic models[4]. Forecasting using neural networks has increased over the past decade and its application to finance has grown rapidly, becoming one of the most explored prediction techniques for stock index returns (Dixit et al., 2013). Its popularity stems from its ability to handle diverse problems in a simplified manner.

Based on the architecture of the human brain, neural networks seek to mimic the brain by using memory and pattern recognition. It is an interconnected group of artificial neurons[5] using mathematical or computational models for information processing. These neurons are structured hierarchically with an input layer, a middle (hidden) layer and an output layer. Various weights are assigned to each connection between neurons which models the impact of an input cell on an output cell. These weights can be either positive or negative, indicating reinforcement or inhibition respectively, of the impact of each input cell on an output cell. In addition, through a training process, the network 'learns' these connection weights through examples from a training set which is presented repeatedly to the network. Each processing element contains an activation

[3] Stochastic models are influenced by other factors, making it a closed system which changes over time.

[4] Deterministic models are a closed system of equations which are independent of being influenced by time.

[5] An artificial neuron is essentially a processor which takes inputs and converts them into outputs to the next neuron based on a particular activation function.

level which may be specified by continuous or discrete values. Activation levels which are determined in response to input signals it receives from the environment are associated with the input layer. Middle and output layers have their activation levels computed as a function of activation levels on the cells connected to it as well as the related connection weights (Malliaris & Salchenberger, 1996).

In order to specify a neural network one would need to select input data, train the network based on a specific algorithm and measure the performance of the network. One of the most essential factors in the formation of a neural network is coming to a decision as to what the network will learn. Input data may vary from raw data on volume or price, or it may incorporate derived data extracted from technical or fundamental indicators. This study attempts to forecast the SAVI using historical values of the index by making use of neural networks, and comparing the forecasting accuracy to more traditional forecasting tools such as various historical volatility models , autoregressive, and heteroskedastic models which are all discussed later.

1.2 Research Problems and Objectives

This study attempts to answer the following two questions from the perspective of an investor wanting to understand volatility and its forecastability in a South African context:

- Can a neural network forecast the volatility of a stock market index?
- Is a neural network a better predictor of volatility than traditional forecasting methods?

In addition, portfolio managers may find it difficult to rebalance portfolios during highly volatile periods. When the ability to forecast volatility becomes clearer, portfolio managers may be able to wait for less volatile periods before rebalancing their portfolios. Furthermore, speculators may want to know when volatility may increase in order to capitalise on short term gains.

As with any study, this study is not without potential issues that may be critiqued. Some issues involved with the implementation of this study relate to the intricacy of the market which may not be adequately captured by a neural network. Secondly, theory suggests that stock markets follow a random walk and that prices cannot be forecasted. With this school of thought in mind,

it may however still be possible to forecast volatility in the stock market.[6] Lastly, a more careful data design may be needed as the type of neural network used may influence the results. After taking into account these potential issues, this study is still feasible and aims to add to the existing plethora of studies which attempt to forecast stock returns, as well as volatility, using both traditional methods, as well as more advanced methods such as neural networks.[7]

1.2 Feasibility of Study

As stated previously, the objective of this study is to forecast the SAVI using neural networks, and to compare a neural network's ability to forecast volatility with more traditional forecasting methods. This study aims to add to the existing body of knowledge on forecasting time-series and at the same time provides a direct decision making support tool for investing on the South African stock market. Other studies have been done using neural networks to predict volatility on the stock market; Samouilhan and Shannon (2008) investigate the comparative ability of three various types of volatility forecasts, but do not include neural networks as a potential forecasting tool. Even though artificial intelligence techniques are fairly easy to use, the end user cannot always decipher how the results from these techniques were computed. Neural networks are sometimes referred to as a "black box.

The data used in this study is fairly common and does not pose a problem. However, the optimal number of hidden layers to be used in the neural network may pose a problem, as there are no hard and fast rules for this. The sample period needs be appropriate to account for any time-specific factors that may have influenced share price performance. Furthermore, the JSE updated the SAVI in 2009, reflecting a new way of measuring the expected volatility, which is more consistent from a theoretical point of view and which now accommodates the way that traders trade options. This may lead to the results having to be interpreted with caution, taking into account the change in the way the SAVI is now calculated.

[6] This idea is discussed further in Chapter 2.1.

[7] The majorities of these studies are done outside of South Africa and are discussed extensively throughout Chapter 2.

1.4 Hypothesis and Theoretical Framework

The following two hypotheses lay the foundation for this study. These two hypotheses have a direct link to each other, but individually are still critical aspects for an investor to consider during the investment decision making process. The theoretical framework is also laid out which provides the basis for the remainder of the study. The hypotheses are as follows:

Null Hypothesis: The SAVI cannot be forecasted using neural networks.

Secondary Null Hypothesis: A neural network is not a better modelling technique than traditional forecasting techniques.

Various computational intelligence techniques have been used to attempt to predict stock prices as well as the volatility in stock markets. Most of the existing literature has neglected emerging markets and the central focus has seemed to be on developed markets, possibly due to the ease of accessing data, as well as the view that developed markets have better functionality. Thus, this study aims to use a neural network to forecast the South African Volatility Index and compare its forecasting ability with more traditional methods of forecasting.

The main purpose of forecasting a stock market is for financial gain and to equip decision makers with information to enable better decision making. Forecasting is the procedure of making statements about future outcomes which are yet to be observed (Gilchrist, 1976). Forecasting methods fall into two categories: qualitative and quantitative methods. Quantitative methods can be further segmented into explanatory and time-series forecasting. Explanatory models are based on the assumption that the variable to be forecast has an explanatory relationship with one or more other variables. When the inputs are changed, the output of the system is affected in a predictable way, assuming that the explanatory relationship remains the same. This is a closed system. On the other hand, time-series forecasting focuses only on information on the variable to be forecast. This study uses time-series forecasting, which enables one to forecast what will happen, without the risk of any spurious relationships arising due to unnecessary variables used as inputs in the forecasting process. A shortfall of using time-series forecasting as opposed to explanatory forecasting is that one can only ascertain what will happen, but not why it happens.

The hypotheses are tested using a neural network and a comparison is made with the more traditional forecasting techniques by means of comparing the RMSE of each forecasting tool. Utilising an artificial intelligence technique can be regarded as an improvement from using other methods such as technical and fundamental analysis in order to predict stock prices, and hence volatility in the market.

Figure 3 below represents the typical architecture of the multilayer perceptron, where the nodes depict the processing unit, the arrows are representative of the connections, and the arrowheads indicate the normal direction of signal flow. A multilayer perceptron is a nonlinear dynamic system[8] which is comprised of many interconnected processing units, called artificial neurons. A multilayer perceptron is able to learn functional relationships from examples and have the ability to discover patterns in data through self-organisation (Seleemah, 2012).

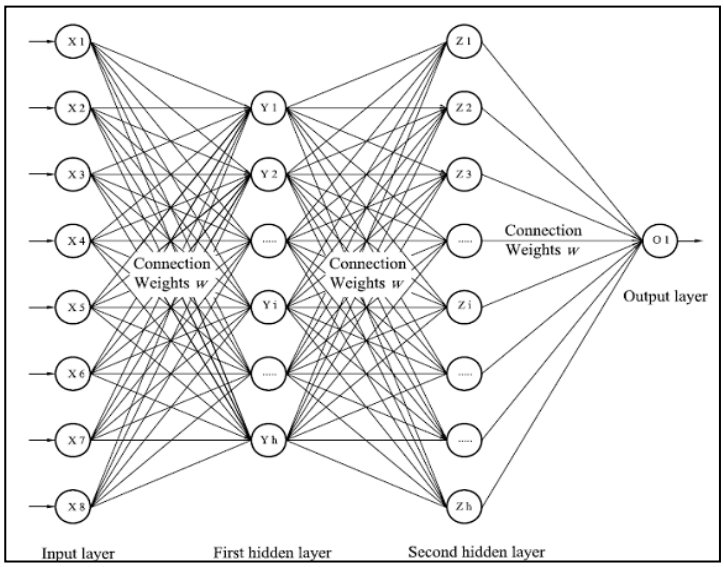


Figure 3: Typical architecture of a multilayer perceptron

[8] This system changes over time. If initial conditions change, then the end result changes, which is also known as a chaotic system.

This study makes use of a Time Delay Neural Network (TDNN), which can be seen as an extension of the aforementioned multilayer perceptron, with time-delayed links (Cancelliere & Gemello, 1996). The definition of a TDNN given by Cancelliere and Gemello (1996, p.33), whereby the authors define the network as "being used to deal with sequence recognition problems in which a finite memory of past events is sufficient," makes this type of network ideal for the time-series investigated in this study.

Figure 4 below shows a network diagram for a simple time delay neural network. There is generally one input node for each feature in one time period of the input sample vector. A condensed set of features arises due to input data to an algorithm being too large. Transforming the input data into the set of features is termed feature extraction (Versace & Wong, 2009). For each added lagging time period incorporated into the network, all features are represented in additional input nodes - similar to adding an additional regressor to an equation. The hidden nodes sum the result of the input layer, where these results are weighted values of the features and are converted using either fast or slow learning techniques. The output nodes sum the result of the hidden layers, which are altered in the hidden layer using a threshold function to produce a Boolean response.[9]

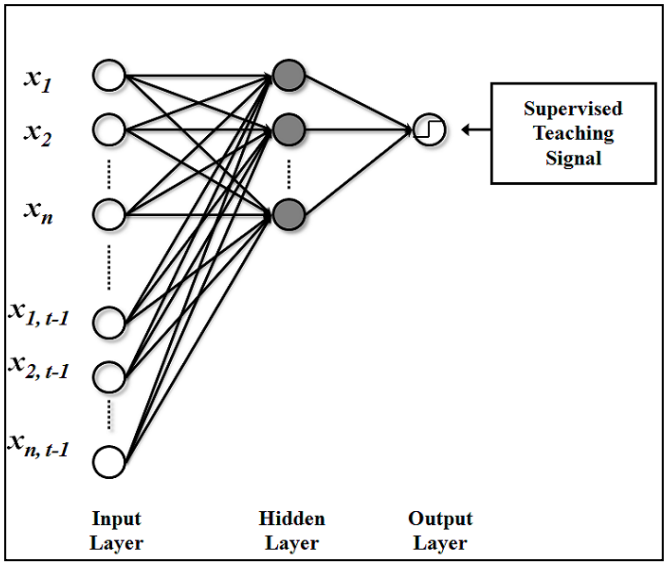


Figure 4: Typical architecture of a time delay neural network (Source: Versace & Wong, 2009, p, 14)

[9] A Boolean response is that which has only two possible types; true, or false.

Figure 5 below shows a flow diagram for creating and testing the model. The first step is collecting the data. This is then followed by processing the data, in order to train the TDNN, after which the model is tested and a RMSE is calculated in order to make a comparison with the other forecasting techniques.

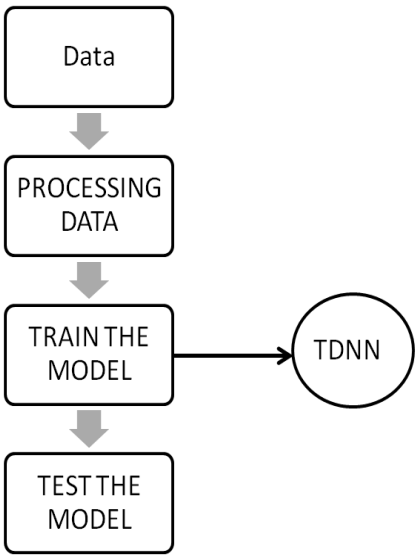


Figure 5: Flow diagram for creating and testing the model

1.5 Summary of Findings

The root mean squared error of each of the forecasting tools is compared, in order to decipher which tool has the best forecasting ability of the volatility in the South African stock market, in particular, the Johannesburg Stock Exchange.

Only historical values of the SAVI are considered and the logarithmic returns are calculated and used as inputs to the time-delay neural network. This method is also used to calculate the return series that is used for the other traditional forecasting methods. Although other inputs could have been used as potential forecasters, the sole use of historical values of the SAVI allows the

neural network to provide a direct forecast, without being affected by possible spurious inputs. Furthermore, literature suggests that forecasting using historical values has sometimes proven to be successful. At the same time, the restricted use of historical values as forecasters leaves open many avenues for further research in this field.

In addition, this study has not been explored extensively in South Africa, giving rise to opportunities for further research. Empirical analysis can be done by utilising different neural networks, with various architectures and neurodynamics. Neurodynamics can be considered as being the foundation of any core computational routines in a neural network. In addition, due to neural networks having the ability to solve complex problems, even when data is convoluted or contains missing values allows one to use additional factors besides historical prices or returns as inputs to the network. In particular, the daily open, close, high, and low values of the South African All Share Index can be used, together with other economic, technical, and fundamental factors as in Lawrence (1997). From a theoretical perspective, one can examine investor sentiment as being a driver of the SAVI, with proxies for this sentiment being used as an additional input in forecasting the SAVI.

This study will proceed as follows: An outline of the literature surrounding the concept of forecasting and its direct relations are examined in Chapter 2. Thereafter, Chapter 3 presents a concise description of the data and methodology used this study. Chapter 4 presents and discusses the results, together with potential biases that were present in this study. Lastly, Chapter 5 considers potential avenues for further research of this study, ending with a conclusion. Appendix A displays various diagnostics and analysis of the process surrounding forecasting using an ARCH model, as well as providing a similar analysis, but with reference made to forecasting using a GARCH model. Appendix B displays the relevant code which can be used to implement testing the TDNN in Matlab™.

2 Literature Review

This chapter begins with a description of the EMH. This discussion will give a definition of an efficient market, and provide evidence of forecastability, or lack thereof in such a market. Thereafter, the literature surrounding traditional forecasting techniques is analysed, with specific reference to linear and non-linear methods. A thorough analysis of the use of artificial intelligence in forecasting is then examined, followed by a discussion on volatility indices, and finally ending with possible drivers of the SAVI which stem from behavioural finance.

2.1 The Efficient Market Hypothesis and Forecasting

According to the proponents of the EMH by Fama (1970), prices fully reflect all available information and investors cannot beat the market by stock picking. This suggests that when this information is fully and freely available, thorough analysis done by many agents are the same, and future changes in share prices cannot be predicted. This notion suggests that predictability is representative of inefficiencies in the way that capital markets function, and that these inefficiencies are exploitable. There is another competing view which argues that predictability is a natural outcome of an efficient capital market (Ferson, 2007). Ferson (2007) claims that the interpretation of predictability, as well as the evidence surrounding it remains controversial, with research on predictability of asset markets being likely to continue and remain both useful and controversial.

It is important to note the difference between ‘forecasting’ and ‘predicting’. Lewis-Beck (2005) defines forecasting as being of a forward looking nature, aiming to tell of events before they happen. The author defines ‘predicting’ as not necessarily having to be forward looking. Furthermore, prediction is said to be part of statistical inference, while transferring knowledge about a subset of a population to the entire population. Information which is transferred across different points can be considered forecasting. Forecasting can be seen as being a subset of prediction, with all forecasts being predictions, but not all predictions being forecasts.

Jensen (1978) states that a market is efficient with respect to information set $\Phi_{t,}$ if one cannot make economic profits by trading on the basis of information set Φ_t. Malkiel (1992) defines a market as efficient if it fully and accurately reflects all significant information in determining

asset prices. The author claims that the market is efficient with respect to some information set Φ_t, if security prices are not affected once this information is revealed to market participants. In addition, the author states that it is impossible to make economic profits by trading on the basis of Φ_t. Granger and Timmermann (2004) state that the EMH gives rise to forecasting tests which parallel those implemented when the optimality of a forecast is tested in the context of a given information set. The authors do however note that there are significant differences which stem from the idea that market efficiency tests rely on establishing profitable trading opportunities in 'real time', whereas forecasters constantly search for predictable patterns, and affect prices when they take advantage of trading opportunities.

It is important to note at this point that this study does not attempt to refute or challenge the EMH, but rather aims to provide an investment tool by which investors may use in order to make better, informed investment decisions, by understanding volatility in the South African stock market through the analysis of the SAVI, as well as through the components by which the SAVI is calculated, together with other factors such as investor sentiment by which the SAVI is defined.[10]

Granger and Timmerman (2004, p.15) state that the EMH is a "backbreaker for forecasters". The authors claim that if a successful prediction model exists, this would immediately be a violation of the EMH. The authors however, also acknowledge the fact that once model uncertainty is taken into consideration, the violation of the EMH is not necessarily valid, unless evidence exists of a search technology that enables investors to find successful models *ex ante.* The authors argue further that forecasters are continuously seeking predictable patterns and have a significant effect on prices when these trading opportunities are acted upon. Moreover, the authors claim that stable forecasting patterns are thus probably not likely to persist for long periods of time and will disappear or "self-destruct" when other investors also become aware of the apparent mispricing.

It is interesting to note that even though Granger and Timmermann (2004) argue that traditional time-series forecasting methods which use individual forecasting models, or stable combinations

[10] The SAVI is also referred to as a "fear gauge" as it is representative of the overall sentiment in the market (Joseph, Koetze & Oosthuizen, 2009).

of these models, are probably not likely to be useful. The authors do claim that simple methods with the ability to adapt or learn quickly may have some forecasting ability. Apart from these models having the ability to adapt or learn quickly, they should also have the ability to be applied to a large number of return series, searching for possible "hot spots" where the capacity to forecast is present (Granger & Timmerman, 2001, p.23) . This study makes use of a time-delay neural network which has the ability to adapt and to learn new patterns quickly from complex and convoluted data, whereby the examination of the forecast errors may indicate such "hot spots" where forecastability is available.

Forecasting in finance, as well as in the field of economics usually involves a variable of interest that is unobservable, even *ex-post*. There is a widely held view that the cornerstone of finance is the trade-off between risk and return. Therefore, measuring and forecasting volatility is argued as being one of the most important pursuits in empirical asset pricing as well as financial risk management. Fama (1991) claims that there is stronger evidence of predictability in returns when using lagged values of returns and information which is publically available. This study uses lagged values of returns as well as information which is freely and publically available. It is acknowledged however, that this is not a complete set of criteria in order for forecastability to be present.

Fama (1970) provides a survey of market efficiency and focuses on testing informational efficiency. The author concludes that there is evidence of weak-form and semi-strong efficiency, whereby weak-form efficiency is described by an information set consisting solely of historical prices, with the information set for semi-strong efficiency consisting of historical prices, as well as information that is publically available. Even with these conclusions drawn by Fama (1970), of markets exhibiting both weak-form and semi-strong efficiency, there are many studies which measure statistically significant serial dependence in daily stock return data (Ferson, 1995). Roll (1984) argues that the serial dependence in daily returns can be explained by fluctuations in end of day price quotes which vary between bid and ask prices. It can be argued that these effects do not however represent predictability and are easily exploitable through any practicable trading strategy (Ferson, 1995). It can be argued that the evidence of predictability in returns proposed by Fama (1991) can also be exploited, and competed away.

In contrast, Lo and MacKinlay (1988) model expected returns as following an autoregressive process which suggests that serial dependence may reflect varying conditional means (see also Conrad and Kaul (1988)). An autoregressive process is formulated around the hypothesis that past values have an impact on current values. Serial dependence can be defined as a value in a time series being statistically dependent on a value at a different point in time (Gilchrist, 1976). If the expected returns follow an autoregressive process, then the actual returns may be described by the summation of an autoregressive process and white noise, resulting in the actual returns following an ARMA process (Conrad & Kaul, 1988). A white noise process can be defined as not having any correlation between its values at different times (Gilchrist, 1976). It should be noted that the aforementioned studies were analysed over short periods of time, with larger sample periods having the potential to alter the stated results.

Forecastability does not necessarily have to be defined over short intervals of time; this study uses daily data, but any investment recommendation should be made with the investment time horizon in mind, together with other critical factors which affect an investor's decision making process. With that being said, DeBondt and Thaler (1985) find evidence that past high-return stocks perform poorly over the following five years, and stocks which yield low returns tend to perform better over the following five years; a conclusion which itself can be considered a form of forecastability. Rozeff (1984) and Berk (1995) put forth an argument that valuation ratios should be a good forecasting variable.[11] Depending on how one defines stock market forecastability, other interesting forecasting variables stem from Merton (1980), who states that sophisticated intertemporal and arbitrage-model versions of the CAPM illustrate that aggregate expected returns on the stock market may depend on additional types of risk, over and above market risk which is the sole determinant of risk in the CAPM, introduced by Sharpe (1964). Lastly, French, Schwert and Stambaugh (1987) come to mixed conclusions by attempting to predict market returns with predetermined market volatility measures. The author's found weak and sometimes negative beta (β) coefficients.

The common misconception that forecasting stock returns is contrary to market efficiency is discussed by Rapach and Zhou (2013). The authors states that the widely used random walk

[11] The idea behind this argument is the relationship that the expected rate of return (r) has with the expected cash flow (c) and share price (P), *where* $r = \frac{c}{P}$.

model which proposes that future stock returns are unpredictable on the basis of currently available information is consistent with market efficiency, but at the same time, so is a forecastable return process – as long as forecastability is exposed to time-varying aggregate risk. This idea revolves around the concept that higher returns are to be matched with higher risks, as proposed by Merton (1980). This in turn has implications for forecasting, whereby only once risk-adjusted time-varying expected returns are nonzero, can one conclude that a market is inefficient (Rapach and Zhou, 2013). Therefore, the ability to forecast stock returns does not necessarily imply that the market in question is inefficient.

It should be reiterated that this study makes no assumption about whether or not the South African market is efficient, while attempting to forecast volatility. This section merely provides a foundation upon which the topic of forecasting in a stock market may evolve. The objectives set out in Chapter 1 are not formulated around the South African market being inefficient, but rather stand as hypotheses in order to ascertain which forecasting tool is a better forecaster of the time-series in question, bearing in mind the conclusions drawn by Bates and Granger (1969) and Granger and Ramanathan (1984), who claim that combinations of forecasts from various models using different information may outperform the forecasts presented by any one particular model.

2.2 Forecasting Volatility Using Historical Data

As mentioned previously, this study uses historical data in order to forecast volatility. A set of historical data may have endless models which may be used to forecast future data. Figlewski (1994) claims that when an asset's price follows a constant volatility lognormal diffusion model, referred to as a Brownian motion equation and represented as:

$$\frac{dS}{S} = \mu dt + \sigma dz \qquad (1)$$

where dS is the asset price change over an infinitesimal time interval, dt, μ is the annual mean return, dz is a time independent random disturbance with a mean of zero and a variance of $1 \times dt$, and σ is the standard deviation (or volatility) of the annual return, then volatility can be easily estimated from historical data.

The price behaviour of the SAVI changes over time and fluctuates over intervals, with these fluctuations varying in length over the different intervals. Figlewski (1994) argues that it is common to compute volatility using historical data as if prices did in fact follow Equation (1). The author states however, that adjustments should be made to the estimation methodology, or the volatility number it produces, in order to counteract any problems which may be suspected to arise.

Standard statistical criteria using historical data are based on assumptions of stability, either through constant variance, or constant parameters of the variance process, which are not likely to hold over extended periods (Figlewski, 1994). If returns are defined as $R = E(R|\Omega) + u$, where Ω is the information at the beginning of the period and u is the unexpected return, then since $E(R|\Omega) = 0$, it is trivial to show that the unexpected return cannot be predicted in advance. This simple result suggests that the forecastability of an efficient market is based on the expected return varying systematically through time (Ferson, 1995). Lo and MacKinlay (1988) provide evidence of returns following an autoregressive process which emphasise the possibility of expected returns varying through time.

Forecasting expected returns is not the only measure of forecastability associated with stock prices. Merton (1980) showed that the mean of a stock return is difficult to predict but that it is immaterial when estimating the conditional variance, when the time between the observation at time *t* and *t+1* is short. Engle (2004) models predictable second moments of returns using ARCH and GARCH-type models. The statistical analysis of the historical data used in this study which are presented later suggests that the second order moment of the historical returns of the SAVI is not constant, but in fact varies through time. Figlewski (1994) argues that using a GARCH model allows volatility to vary systematically over time, but the GARCH parameters themselves must be constant.

2.3 Forecasting Volatility using Traditional Methods

Investors do not always have access to sophisticated forecasting tools like neural networks. In addition, neural networks and its application to finance have not always been in the forefront of financial literature. Although traditional forecasting methods may be seen as a naive approach to

forecasting volatility, the plethora of studies surrounding the topic advocates that it is still an area of interest to professionals throughout the world. The existing literature suggests that no one forecasting tool has successfully been deemed as being superior to another consistently enough to draw a conclusion about the best forecasting method. As the debate surrounding market efficiency continues to expand and accept new innovative possibilities with which to explore the topic, so shall the debate around forecasting and which tool has the best forecasting ability, as these concepts are in fact all interrelated.

Arguably, the simplest forecasting result stems from the random walk hypothesis, which is based on the assumption that the best prediction of tomorrow's price is the current market price. The random walk hypothesis was found to predict the direction of the market 50% of the time, whereas linear regression models were found to predict future prices about 55% to 65% of the time (Campbell, 1995).

This study does not explore any linear regression models and their forecasting ability, but does consider a historical average model as one of the simpler forecasting tools. This method of evaluating volatility can be seen as a naive approach due its simplicity, and the assumption that future volatility is in fact the historical average. Evidence suggests that the conditional expectation of volatility is time-varying (Bollerslev, Chou & Kroner, 1992). It is for these reasons that the historical average model and the literature which surrounds it is not explored extensively; its volatility estimate is stated merely as a basis with which to compare the other, more advanced methods.

It is almost inevitable, and no secret, that all nations will at some point experience advanced social, economic or political changes which make forecasting with constant parameter linear models particularly difficult (Marcellino, 2007). Campbell (1995) however, claims that linear regression forecasting models have been valuable in predicting returns in both developed and emerging markets. A slightly different result was reached by Geweke and Meese (1984) who examined various linear models and did a comparison using 150 macroeconomic time-series, concluding that Autoregressive (AR) models performed well.

It was Yule (1927) however who initiated the paradigm of stochasticity in time-series by hypothesising that all time-series can be considered a realisation of a stochastic process. The

author also contributed to the concept of autoregressive (AR) and moving average models (MA). Brock, Lakonishok and LeBaron (1992) show that MA technical trading rules have some predictive abilities in determining both first and second order moments over the entire ninety year period that was examined. On the other hand, Sullivan, Timmernan and White (1999) argue that although these rules show little or no evidence of *data snooping*, the forecasting performance of these rules seems to have disappeared over time. "Data snooping occurs when a given set of data is used more than once for purposes of inference or model selection" (White, 2000, p.1097). This study does not directly make use of technical indicators in the forecasting process, and is left for future research.

An additional linear forecasting technique is that of exponential smoothing, whereby more recent observations are given more importance and thus considered better forecasters. There are many types of exponential smoothing techniques, with this study making use of Simple Exponential Smoothing (SES), which is explored later. Exponential smoothing did not receive a lot of attention from statisticians over the past twenty-five years (See De Gooijer and Hyndman, 2006). After a statistical basis was provided for exponential smoothing methods, Gardner (1985) later achieved satisfactory results by introducing exponential smoothing methods into supply chain management, in order to predict demand. Chen, Daouk and Leung (2000) used an adaptive exponential smoothing model, which takes seasonal variations into account by means of a coefficient which is allowed to fluctuate through time in order to detect any changes in the event being studied, to forecast the Nikkei 225 indices. The authors' empirical analysis suggests superior forecastability.

Muth (1960) was the first to advocate a statistical basis for SES. The author showed that optimal forecasts for a random walk plus noise could be obtained by making use of SES. Box and Jenkins (1970) and Roberts (1982) contributed towards placing exponential smoothing within a statistical framework and found that certain exponential smoothing forecasts occur as unique cases of ARIMA models. Satchell and Timmermann (1995) provide evidence of SES being advantageous for a wide range of data generating processes, and Hyndmann (2001) argues that SES performs better than first order ARIMA models. The author claims that SES does not have as many model selection problems, particularly when data is non-normal. Anderson, Carbone,

Fildes, Hibon et al. (1982) examined the performance of univariate methods in many time-series and concluded that ES was often successful.

Miller and Williams (1999) provided modifications to exponential smoothing models, in order to deal with discontinuities. Further variations of the original methods arise from Guerrero and Rosas (1994), who examined exponential smoothing forecasts conditional upon one or more constraints. The existing literature boasts many more variations, with Archibald and Koehler (2003) developing a new system of renormalisation equations which are more straightforward, yet give the same point forecasts as the original methods.

Another forecasting tool which is explored in this study is the Exponentially Weighted Moving Average (EWMA) estimator, which has proven to be successful at forecasting the volatility of returns over short horizons. Stuart (1986) posits that the EWMA was born from the early work of econometricians, and although its use has been recognised, it still remains a neglected tool. There is also evidence of this method outperforming more sophisticated forecasting methods such as GARCH models (see Boudoukh, Richardson and Whitelaw, 1997 and Alexander and Leigh, 1997). Further evidence of EWMA models being successful predictors arise from Tse (1991) and Tse and Tung (1992), who claim that EWMA models provide more accurate forecasts than GARCH models by using data sets from Japanese and Singaporean markets respectively. Ladokhin (2009) compared various forecasting tools, testing the accuracy of the models by attempting to forecast the S&P 500 stock index. The author found that the EWMA estimator produced better forecasting ability than the historical average, exponential smoothing estimate, all ARMA, ARCH and GARCH models examined, as well as the two neural networks which were also compared. The author used a neural network that was a generalisation of the networks prescribed in the literature. It is possible that different neural networks have different forecasting abilities, and the decision of which neural network to use should be made with careful analysis of the problem at hand. A TDNN could have resulted in improved forecasting ability as opposed to utilising a generalised network. The results are based purely on the analysis of the RMSE, where the lower the value, the better the predictive ability, *ceteris paribus*. This study adopts a similar methodology of comparison and is discussed later.

Many people believe that linear models are ‘a relatively poor’ method of forecasting certain types of economic behaviour. The possibility that the effects of shocks accumulate before they

‘explode’, as well as the idea that certain variables may only be forecasted once in a while hints at the requirement of non-linear forecasting techniques (Clements, Franses & Swanson, 2004). Moreover, it is conceivable that complicated forces which drive economic events initiate nonlinear dynamics into aggregate time-series variables, which in turn means that forecasting would involve identifying and exploiting these nonlinearities (Granger, 1993).

The complex nature of markets, as well as the discovery of markets exhibiting non-linear movements (see Abhyankar, Copeland and Wong, 1997) may be reasons as to why linear and historical models do not distinctively outperform random walk models in their predictive abilities. Nelson (1992) found evidence of the ARCH model being able to perform well for volatility forecasting when using high frequency data, even when the model is misspecified. Ladokhin (2009) found the ARCH model to have poor predictive ability, with the ARCH model only displaying better performance than the simple historical average, which itself is not known to have the best predictive ability due to its naive simplicity.

Samouilhan and Shannon (2008) investigate the comparative ability of ARCH, implied volatility forecasts and other historical average models. The authors find that simple ARCH models provide a good in-sample forecast, but are however the worst predictors of volatility, together with Historical Volatility Models (HIS) models. The authors argue that more complex ARCH and GARCH models are the best models to use to forecast volatility in South Africa.

A more parsimonious model is the generalised ARCH (GARCH) model (Taylor, 1987), where additional dependencies are permitted on lags of the conditional variance. Akigray (1989) was one of the earliest authors to examine the predictive power of GARCH models. The author found that GARCH models consistently outperform historical average and EWMA models in all sub-periods and evaluation measures. Sabbatini and Linton (1998) provide evidence of the simple GARCH (1,1) model providing a good parameterisation for the daily returns of the Swiss market index.

Another method of forecasting or extracting volatility from a stock index is via the Black-Scholes option pricing model (Ladokhin, 2009). The model allows one to compute a volatility forecast, if the option market is efficient and the valuation model is correctly specified. Furthermore, this implied volatility forecast is market-determined, and is only possible if all

other option pricing parameters are "objectively" available (Fleming, 1998, p.318). Beckers (1981) found that the implied volatility does in fact contain information about future volatility. The author did however use a fairly small dataset, but later studies such as Day and Lewis (1992) and Lamoureux and Lastrapes (1993) who used S&P 100 Index options and individual equity options respectively, found that implied volatility does contain useful information in forecasting volatility, but that time-series models contain more information about future volatility and may be better forecasters.

This study does not look at implied volatility as a forecasting measure of future volatility, as measurement error is problematic for index options as they are not always traded frequently, giving rise to infrequent trading effects on the underlying index price. Further measurement error may arise from the mismatch between the times that stock and option markets close (Fleming, 1998). Furthermore, Fleming (1998) states that implied volatility is not covariance stationary and is almost a unit root process. Lastly, this study focuses on forecasting the SAVI, an index which will not be able to be priced by using a standard option pricing model. The pricing of the SAVI is explained later.

2.4 Forecasting Volatility using Artificial Intelligence

Forecasting using neural networks is not new and the existing body of knowledge contains a vast amount of literature which compares neural networks to traditional historical techniques for forecasting. Zhang (2001) claims that neural networks are successful in linear time-series modelling and forecasting. This seems to suggest that artificial intelligence techniques such as neural networks can compete with linear models of forecasting. An important point to note is that unlike the forecasting methods discussed earlier, neural networks are data-driven, self–adaptive methods, whereby there are only a few *a priori* assumptions about the models used (Hu, Patuwo & Zhang, 1998). In addition, Hu, Patuwo and Zhang (1998) claim that neural networks are highly suited for problems whereby the solutions require knowledge that is not easily specified, but where there are enough data and observations. This implies that neural networks learn from examples, and detect functional relationships in the data which are difficult to describe (Hu, Patuwo and Zhang, 1998).

Werbos (1974) first described the process of training artificial neural networks through the backpropagation of errors and concluded that neural networks trained with backpropagation outperform traditional statistical forecasting methods, including regression and Box-Jenkins approaches. The concept and process of a backpropagation algorithm is explored later. Amiri, Von Rossen and Zwanzig (2009) describe Box-Jenkins approaches to time series modelling as consisting of a systematic class of models called ARIMA. These models are said to be multiplicative models, which means that an assumption is made that observed data results from the products of factors involving various operators which respond to a white noise input.

Literature suggests that traditional methods of forecasting have variations in the way models are utilised or combined. This is also true for the countless collection of artificial intelligence techniques. Donaldson and Kamstra (1997) make use of a Neural Network-GARCH model in an attempt to capture volatility effects of stock returns. The authors find that both in-sample and out-of-sample comparisons suggest that their neural network model captures certain volatility effects which are overlooked by GARCH models. Dockner, Dorffner and Schittenkopf (2000) conclude that volatility predictions from neural networks are superior to GARCH models. On the other hand, Gahan, Mantri and Nayak (2010) and Anwar and Mikami (2011) argue that ARCH (GARCH) models are superior to a neural network model.

Bollerslev et al., (1992) argue that, with but a few exceptions, the majority of research into volatility utilises data from the United States (US), the United Kingdom and Japanese markets. This study thus examines data from South Africa, in particular the Johannesburg Stock Exchange (JSE). The JSE is open and liquid, but at the same time displays characteristics which are different to those of developed markets' bourses.

2.5 The South African Volatility Index – An Introduction

The Chicago Board Options Exchange (CBOE) introduced a volatility index (VIX) in 1993 which was based on the calculations of the S&P 100 stock index options. The VIX has since been re-evaluated in terms of the way that it is calculated, as well as what serves as its core component in assisting traders and hedgers with more accurate figures of stock market volatility.

Furthermore, the VIX provides investors with a tool in order to examine the changes in the volatility of the stock market, and has also become a leading indicator for forecasting and determining the performance of stock markets in the US (Liu & Yang, 2012).

The higher the price of the VIX, the higher is the expected volatility in the stock index, whereas a lower VIX price reflects only a modest expectation of stock index fluctuation. The VIX has been termed the "investor fear gauge" as it is believed to reflect investors' expectations (Liu & Yang, 2012, p.217).

The VIX quickly reaching a level over 40 during the "WorldCom" bankruptcy in 1998, and even exceeding a value of 80 when the Lehman Brothers filed for bankruptcy in 2007, suggests that a volatility index is a good indicator of market information in security markets (Liu & Yang, 2012).

A similar index was developed in South Africa, called the South African Volatility Index (SAVI). The SAVI can be seen as a forecast of equity market risk in South Africa and was first introduced in 1997. The SAVI very swiftly became the benchmark for measuring market sentiment, and is now also referred to as a "fear gauge" – this is primarily due to the negative correlation which is present between the underlying index level and it's volatility, as depicted in Figure 3 below. Figure 3 shows that as the ALSI40 falls, the volatility rises. It is within this context that volatility is seen as a "fear gauge".

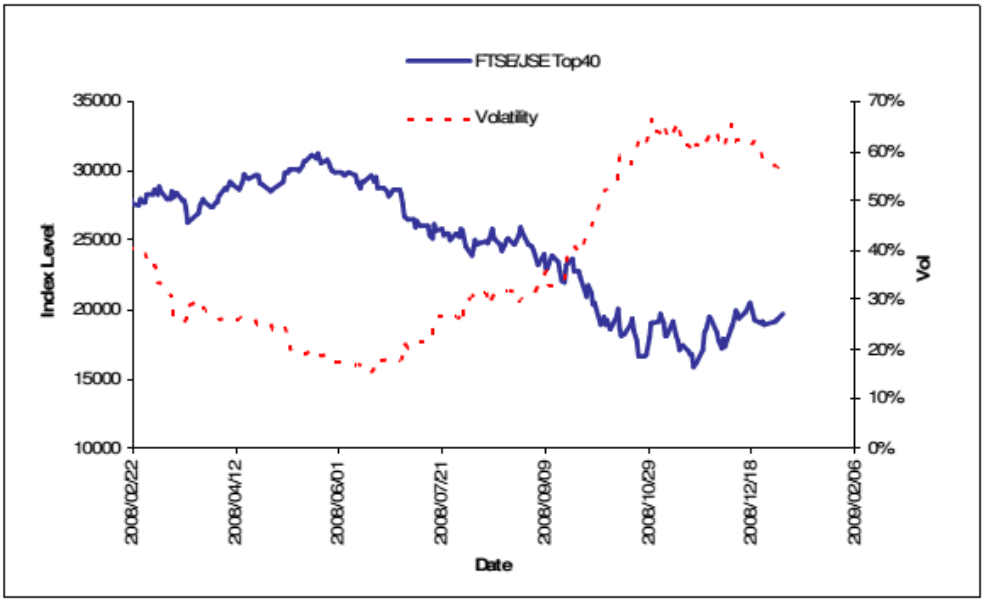


Figure 6: FTSE/JSE Top40 index level and its volatility.

The JSE updated the SAVI in order to reflect a new method of capturing expected volatility. The reason for the revision of the SAVI calculation was to align the SAVI with the theoretical framework and technique that traders use when trading options. The SAVI was previously calculated daily, by means of polling in the market. The polled at-the-money volatilities were used to determine the three-month at-the-money volatility, with the average published as the SAVI (Joseph, Koetze & Oosthuizen, 2009).

The SAVI is now not a polled volatility measurement. This reduces the probability that the calculated volatility can be manipulated by polled volatility contributors. The SAVI is now calculated as the weighted average prices of calls and puts across a diverse array of strike prices that expire within the following three months. In addition, volatility skew is incorporated in the calculation, which reflects the market's expectation of a market crash (Joseph, Koetze & Oosthuizen, 2009). The volatility skew is a function relating the implied volatility of an option to its strike price. Furthermore, in order to ascertain whether or not the observed volatility skew is fair, a derivation of skew should to be made which is independent of model assumptions and subjective market expectations (Araujo & Mare, 2006).

Figure 4 below depicts the new SAVI in blue, with the old SAVI in red. It can be seen that the incorporation of the volatility skew in the new SAVI calculation makes the new SAVI slightly different to the old SAVI.

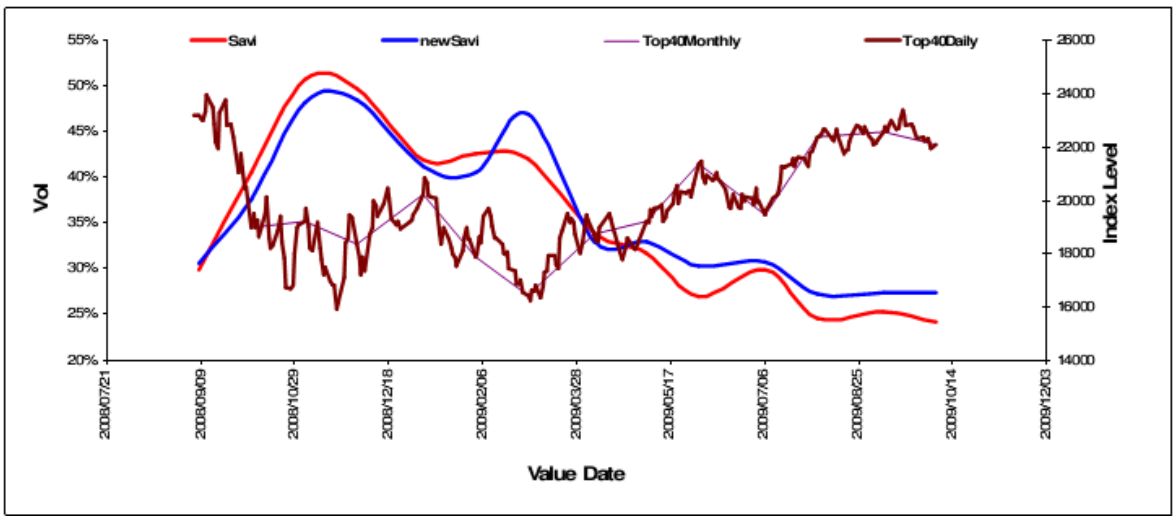


Figure 7: The SAVI and the new SAVI. The FTSE/JSE Top40 index level is also plotted. (Source: Joseph, Koetze and Oosthuizen (2009))

As stated by the Joseph, Koetze and Oosthuizen (2009), the SAVI is calculated by means of the following mathematical equation:

$$SAVI = \sqrt{\sum_{i=1}^{n=F} w_{ip} P_i(K_i) + \sum_{i=n}^{\infty} w_{iC} C_i(K_i)} \quad (2)$$

where F is the current forward of the FTSE/JSE Top40 index level, which is obtained using the risk free interest rate and dividend yield. $P_i(K_i)$ are the liquid put options and $C_i(K_i)$ are the liquid call options, with each option having a strike price K_i. The put and call options are priced using the traded market volatility skew which expires in three month's time. Joseph, Koetze and Oosthuizen (2009) claim that calls and puts may be used to find the price of volatility, given that option prices are directly proportional to their input volatility.

Furthermore, the three month (T) volatility skew, $\sigma_{K(0,T)}$, is defined using the time weighted interpolation function:

$$\sigma_{K(0,T)=} = \sqrt{\{T_2 \sigma_K^2(0,T_2)\left[\frac{N_1}{N_2 - N_1}\right] + T_1 \sigma_K^2(0,T_1)\left[\frac{N_2}{N_2 - N_1}\right]\}\left[\frac{N_0}{N_3}\right]} \quad (3)$$

where N_1 is days to the near skew from the three month skew expiry date, and N_2 is the next nearest skew from the three month skew expiry date. N_0 is the number of days in the year[12], and N_3 is the number of days from the value date to the three month date (Joseph, Koetze & Oosthuizen, 2009).

Araujo and Mare (2006) state that stock market participants view volatility skew as a means by which market makers may extract more profitability from option trades. The authors further note that a bid-offer spread is incorporated in implied volatility quotes, which reflects potential profit to the market maker, while the skew reflects potential risk, and should not be interpreted as a risk margin. The volatility skew measure is important in the South African market whereby some

[12] 365 is the South African convention.

options are illiquid, with prices not always being transparent (Araujo & Mare, 2006). In addition, volatility skew is an important component in the way the new SAVI is calculated.

2.5.1 The Benefits of the new SAVI

As expressed in Equation (2), the closer an option is to being at-the-money[13], the more heavily weighted it is in the SAVI calculation, and therefore regarded as more important than far out-the-money volatilities. It is important to note however, that these far out-the-money volatilities are not insignificant in pricing the SAVI, as it is these options that define the volatility skew which is a significant factor in differentiating the new SAVI from the way in which it was previously calculated (Joseph, Koetze & Oosthuizen, 2009). The authors state that by incorporating the volatility skew in the calculation of the SAVI, all dimensions of volatility are considered. In particular, volatilities do not only depend on the time dimension, but also the strike level dimension. Moreover, the authors claim that the SAVI is not model dependent, as it is a weighted average of traded option prices, and therefore does not have to explicitly utilise models such as the Black-Scholes model to derive implied volatility.

2.6 Summary

This chapter covered various aspects of forecasting, including the forecasting of volatility in stock markets. The discussion began with the introduction of market efficiency, together with various studies dedicated to forecasting led to the discussion of forecasting using historical data, as well as forecasting using linear methods. The consensus is mixed, and the literature seems to suggest that no single method of forecasting can be given dominance over any other. These results called for more sophisticated methods of forecasting which may be able to deal with further complexities in markets, some of which have been identified as *volatility clustering*, and *nonlinearity*. This led to the discussion of nonlinear forecasting methods of which still no

[13] An option is said to be at-the-money when the market price of the underlying asset and the option's strike price are equal.

consensus seems to have been reached surrounding which forecasting technique produces the most accurate results. A discussion of artificial intelligence techniques used as forecasting tools was then presented, with the conclusion that these techniques have shown to exhibit superior forecasting ability under certain circumstances, but which artificial intelligence technique to use is still highly debatable. Furthermore, the SAVI was introduced, with specific reference to the way in which it is calculated. Neural networks seem to be a good answer to some of the challenges presented in the literature surrounding forecasting and the traditional framework for examining returns and its volatility.

3. Research Methodology

This chapter commences with an outline of the data and sample used, followed by a description of the methodology used for each forecasting technique. The existing body of knowledge provides evidence of making use of more than one forecasting tool, in an attempt to provide superior forecasting results. The various forecasting tools presented in this study are all interrelated, but the explicit combination of these tools as a forecasting method is beyond the scope of this study and is therefore omitted. A typical forecasting strategy would make use of a series of returns of the index to be forecasted, spanning the sample period. This series is often split in order to evaluate the results of a forecast. This study adopts a similar approach, with the various methods discussed below.

3.1 Data

Data is collected from McGregor BFA. The data consisted of the daily price levels of the South African Volatility Index (SAVI) over the period February 2007 to December 2013, resulting in 1713 price levels and 1712 input logarithmic returns. It is important to note that the SAVI was only introduced by the Johannesburg Stock Exchange Ltd. (JSE) in 2007, to measure the market's expectation of the three-month implied market volatility, with the SAVI then being

updated in 2010 with an improved method of calculation[14]. Furthermore, the SAVI is published at the close of business each day by the South African Futures Exchange (SAFEX), and thus leaves the published opening and closing price levels unchanged throughout the day, together with the high and low price levels. This makes relying solely on the historical daily returns of the SAVI for each forecasting methodology practical and efficient, in particular, when implementing the neural network. Lastly, the logarithmic returns are calculated from the daily prices and are consistent throughout the study. Using logarithmic returns of the SAVI makes sense, as the SAVI is based on the FTSE/JSE Top40 index level. Logarithmic returns take into consideration that stock prices cannot be less than zero, thus creating a lower bound of zero, and creating the assumption of lognormal returns.

Most previous studies comparing forecasting techniques in South Africa have not included neural networks as a potential forecasting tool. This study takes a simple approach with only one input variable; that of historical index data (time-series forecasting). Furthermore, various standard statistical tests for normality and stationarity are done in order to 'clean' the data, before using it in the neural network. These results are presented in Chapter 4.

3.2 Historical Volatility Models

Historical volatility models (HIS) are regarded as some of the most straightforward models used to forecast volatility. It is important to note that these models differ from implied volatility models which calculate volatility such that the theoretical price is equal to the market price of an option.[15] HIS models may vary according to the number of lag volatility terms, as well as the weights assigned to them. This is indicative of the choice on the trade-off between increasing the magnitude of information and using more recent information. The following HIS models are used in this study.

[14] The SAVI is explored extensively in Chapter 2.5.

[15] This can be done by solving for the volatility variable, through an iterative process, using the Black-Scholes option pricing formula.

3.2.1 Historical Average Model (HAM)

The HAM makes forecasts based on the entire history of volatility as opposed to the Random Walk model which uses today's volatility as the best forecast for tomorrow's volatility. The HAM is calculated by using the following formula:

$$\hat{\sigma}_{t+1} = = \frac{1}{t}\sum_{i=1}^{t}\sigma_i \quad (4)$$

where $\hat{\sigma}_{t+1}$is the next period's standard deviation, which is used as a measure of volatility.

3.2.2 Simple Exponential Smoothing (SES)

ES is a type of adaptive forecasting, which is useful when there are only a few observations with which to base a forecast. Furthermore, the forecasts from exponential smoothing methods do not use fixed coefficients like forecasts from regression models, but are adjusted according to previous forecast errors (Bowerman & O'Connell, 1979).

ES can be used to forecast volatility based on historical values. The ES method is described by the following formula:

$$\sigma_t = (1-\alpha)\sigma_{t-1} + \alpha\hat{\sigma}_{t-1} + \zeta_t \quad (5)$$

$$\hat{\sigma}_{t+1} = (1-\alpha)\sigma_t + \alpha\hat{\sigma}_t \quad (6)$$

where α is a smoothing parameter which is estimated by minimising the in-sample forecast errors, ζ_t, as proposed by Poon (2008).

This method is similar to the HAM but more weight is assigned to the recent past and less weight to the more distant values. As discussed in Chapter 2, exponential smoothing has been found to be an effective forecasting method by many researchers. However, ES is a segment of linear forecasting methods and thus is not able to capture nonlinear features of financial time-series.

3.2.3 Exponentially Weighted Moving Average (EWMA)

The formula for determining the moving average with exponential weights is presented by Ladokhin (2009) as follows:

$$\hat{\sigma}_{t+1} = \frac{\sum_{i=0}^{\tau-1} \alpha^i \sigma_{t-1}}{\sum_{i=0}^{\tau-1} \alpha^i} \quad (7)$$

where the smoothing parameter α is calculated as before, by minimising the error on the training set. Shumway and Stoffer (2011) argue that the ARIMA (0,1,1) model leads to an EWMA model. Furthermore, the authors define a smoothing parameter, which is bound between zero and one, whereby the smaller the value of the smoothing parameter, the smoother the forecasts. The authors claim that forecasting with EWMA is popular due to its ease of use, and the need to only retain the previous forecast value and the current observation in order to forecast the next time period. On the other hand, the authors refer to the model as being "abused", as the value of λ, which is used to define the smoothing parameter, is arbitrarily picked by the forecaster.

This study follows Shumway and Stoffer (2011) in order to forecast using EWMA, by adapting the ES model using the HoltWinters method, setting the parameter for α equal to $1-\lambda$, and the parameters β and Φ equal to zero. The parameter, λ, is described later, and parameters β and Φ stem from the HoltWinters three parameter model. The HoltWinters method allows for a series to be modelled with a linear time trend, with additional seasonal variation. It should be noted that setting parameter Φ equal to zero is not equivalent to simply using a two parameter HoltWinters model, as setting Φ to zero only restricts any seasonal factors from changing through time so that nonzero seasonal factors still remain in the forecasts.

3.3 Autoregressive and Heteroskedastic Models

3.3.1 Autoregressive Moving Average (ARMA)

The ARMA model combines both the simple moving average model and an AR model. The ARMA model is obtained from the simple regression method of forecasting, which depicts volatility as a function of its past values and an error term, but allows past volatility errors to also be included. The following formula shows the two components of the ARMA model:

$$\hat{\sigma}_{t+1} = \sum_{i=1}^{p} \lambda_i \sigma_{t+1-i} + \sum_{i=1}^{q} \gamma_i \zeta_{t+1-i} \quad (8)$$

where λ_i and γ_i are parameters of the model, and ζ_1, ζ_2 ... ζ_t are the error terms of the model $\sigma_i = \hat{\sigma} + \zeta_i$. The parameters, length of the moving average,p, as well as the autoregressive term, q, are found by minimising the error on the training set (Ladokhin, 2009)

3.3.2 Autoregressive Conditional Heteroskedasticity (ARCH)

The ARCH model used for forecasting time-series is a non-linear model which does not assume that the variance is constant. Volatility clustering or volatility pooling is a motivating factor for using ARCH models to forecast volatility. Mandelbrot (1963) and Fama (1965) both claim that financial markets display volatility clustering. The ARCH model is described by the following formulae as presented by Ladokhin (2009):

$$r_t = \mu + \varepsilon_t \quad (9)$$

$$\varepsilon_t = \sqrt{h_t z_t} \quad (10)$$

$$h_t = \bar{\omega} + \sum_{j=1}^{q} \alpha_j \varepsilon_{t-j}^2 \quad (11)$$

where r_t is the return at time *t*, μ is the mean return, ε_t are the residuals, $\bar{\omega}$ and α_j are parameters of the model, h_t is the conditional variance with $z_t \sim iid$ N(0,1) normally distributed random variable. The process z_t is scaled by h_t which follows an autoregressive process. In order to ensure that variance h_t is positive, $\bar{\omega}>0$ and $\alpha_j \geq 0$ (Ladhokhin, 2009).

3.3.3 Generalised Autoregressive Conditional Heteroskedasticity (GARCH)

The GARCH model was first developed by Bollerslev (1986). The GARCH model differs from the ARCH model by the form of h_t, which is described as:

$$h_t = \bar{\omega} + \sum_{i=1}^{p} \beta_i h_{t-i} + \sum_{j=1}^{q} \alpha_j \varepsilon_{t-j}^2 \quad (12)$$

where the parameter $\beta_i>0$.

Hamilton and Susmel (1994) argue that a "shock" to a stock market on a given week may have effects on the variance of the market more than a year later. Furthermore, the authors state that these effects are *not* negligible. This may leave GARCH volatility forecasts being seen as too smooth and too high across periods with significant varying levels of instability (Marucci, 2005).

3.4 Neural Network Model Construction

There are a vast number of ways in which a neural network may be constructed. The term "neurodynamics" may be used to describe the properties of an individual neuron such as its transfer function and how the inputs are combined, with the term "architecture" being used to define a neural network's structure in terms of the number of neurons in each layer as well as the number and type of connections.

Boyd and Kaastra (1996) state that determining the number of input neurons is one of the easiest parameters to define once the independent variables have been pre-processed, due to each independent variable being characterised by its own input neuron.

3.4.1 Classification versus Regression

There is a vast array of studies done on predicting the future movement of stock prices using artificial intelligence techniques (Kim & Shin, 2007). Within these artificial intelligence techniques lies the classification problem, which differs from forecasting by means of a regression. Kim and Shin (2007) define the classification problem as estimating a function $f: R_N(-1,1)$ based on input-output training data which is generated from an independently, identically distributed ($i.i.d$) unknown probability distribution, $P(x, y)$, such that f has the ability to classify previously unseen (x, y) pairs. In the case of a stock market index, which is similar to a volatility index, an output of 1 would represent a future price or index value increase, whereas a value of -1 or 0 would represent a decrease. This information can then be used by investors as a tool for investing, assisting investors on whether to buy or sell securities being analysed by means of both the movement of the stock index, as well as the volatility index which reflects investor sentiment as discussed previously.

Forecasting results by means of classification has its flaws however, as the predicted increase or decrease does not inform the investor by how much the stock or volatility index has moved, but merely signals the direction in which it has moved. It is therefore difficult to make accurate decisions of whether to buy or sell a share without knowing the margin or change.

Regression analysis on the other hand involves predicting raw price values. This study does not make use of regression analysis *per se*, but does examine traditional forecasting models which can be used to derive forward looking price levels. A function which allows one to forecast price level at time *t+1* can be stated as follows:

$$P_{t+1=} \int(\mathrm{P_{t,}P_{t-1}}, \dots, \mathrm{P_{t-n}}) \tag{13}$$

where P_{t+1} is the predicted price in the future at time *t+1*, with $P_{t-1}, \dots, P_{t-n}$ representing all previous prices, with P_t being the current price of the index.

3.4.2 Multilayer Perceptron (MLP)

MLPs are feedforward neural networks which are trained with the popular and effective back propagation algorithm. A feedforward neural network consists of a number of layers whereby information moves in one direction only (forward), from the input nodes (Pissareknko, 2002). MLPs require a desired output in order for the supervised network to be able to learn. A MLP consists of at least three layers: the input layer, one or multiple hidden layers and an output layer. The individual nodes are connected by links, each having a certain weight. Each node accepts several values as inputs, which are then processed to produce an output, which can then be 'forwarded' to other nodes. According to Pissarenko (2002), for any given node, *j*, its output is equal to

$$Oj = transfer\ \sum(x_{ji}w_{ji}) \tag{14}$$

where Oj is the output of node *j*, x_{ji} is the i^{th} input to unit *j*, w_{ji} is the weight related to the i^{th} input to unit *j* and *transfer* is a transfer function. A neuron may have many inputs but only one output.

3.4.3 Time Delay Neural Network (TDNN)

Principally, a Time-Delay Neural Network is an extended MLP. TDNNs apply time delays on connections, which allow the neural network to have a "memory", in order to deal with various time series forecasts. This type of NN specifically addresses the time series dependence of data on preceding values. A TDNN has the ability to allow inputs to arrive at hidden units at different points in time, thus allowing the various inputs to be stored for a long enough period of time to have a significant influence on subsequent inputs. The output pattern at a specific point in time is a function of the inputs for that time, as well as the inputs for a prior number of time periods. TDNNs are said to function like a moving average regression model or a finite impulse response filter (Juang & Rabiner, 1993).

Effectively, by the formulation of T time delays, Δt, every neuron has access to each input value at $T+1$ different points in time. The neurons in the neural network can therefore identify relationships between current and previous input values. Furthermore, the network is able to estimate functions that take prior input signals into account[16] (Kaiser, 1994).

Kaiser (1994) posits that traditional methods which are used to speed up backpropagation learning can also be applied to the TDNN. Furthermore, the author states that delayed or scaled input signals can be dealt with by utilising the original definition of the TDNN, which requires all links of a neuron which are coupled to one input to be identical.[17]

The benefits of using a MLP and TDNN are the same as the benefits which underscore the use of a NN, as discussed earlier. TDNN are however difficult to implement due to the large number of input nodes. This study however, takes a simple approach in forecasting the SAVI, by ignoring external inputs which affect the movement of the index.

3.4.4 Learning Environment: Variable Selection

One of the most critical steps in formulating a neural network is deciding on what information to use to teach it; knowing which variables are essential to the market being forecasted is crucial.

[16] Such functions may depend on the derivative of the input.

[17] For control purposes, current and past inputs may be assigned individual weights (Kaiser, 1994).

The main objective of most financial networks is to decide when to buy, sell or hold shares, based on past market indicators. Economic theory can play a vital role in selecting variables which are likely predictors. These predictors are usually used as inputs to the neural network. A decision would need to be made of whether to incorporate both technical and fundamental economic inputs from one or multiple markets.

Boyd and Kaastra (1996) define technical inputs as being lagged values of the dependent variable, or indicators calculated from the lagged values. The authors define fundamental inputs as economic variables which may influence the dependent variable. The author's posit that the most straightforward neural network uses lagged values of the dependent variable(s), or its first difference as inputs. Deboeck (1994) however, suggests that a more popular approach is to compute technical indicators which are based solely on historical prices of the market being forecasted. The interconnectedness of domestic markets and the direct link these markets have to the global economy may allow one to further improve forecasting performance by utilising inter-market data such as cross exchange rates and interest rate differentials (Boyd & Kaastra, 1996). Lawrence (1997) uses a system based on sixty-three indicators from various segments of the economic and financial database.

The frequency of the data used as inputs in the neural network is itself an important factor to take into consideration. Boyd and Kaastra (1996) hypothesise that investors with a long term view may opt to use weekly or even monthly data as inputs to the neural network, in order to formulate the most attractive portfolio of assets, as opposed to using a passive buy and hold strategy. Furthermore, the authors state that an economist interested in forecasting the gross domestic product (GDP), unemployment, or any other economic indicator may prefer to use monthly or quarterly data. On the other hand, Crone and Kourentzes (2008) state that higher frequency data can provide additional detail, which assists in the formation of better forecasts. The authors do however note that because of the higher frequency data, more sophisticated methods are needed in order to analyse outliers. Outliers in high frequency data tend to persist for several observations (Crone & Kourentzes, 2008).

3.4.5 Data Pre-processing

Data pre-processing is usually a crucial step in the implementation of a neural network, in order to achieve better prediction performance. Atiya, El-Sherif, El-Shoura and Shaheen (1999, p.403) state that preprocessing can be considered an "art" with no set of rules defining it. The authors describe input and output preprocessing as a means of extracting features from the inputs and converting the target outputs in such a way that it is easier for the network to extract useful information from the inputs, and provide links to the required output. The authors go so far as to claim that the inputs selected together with the preprocessing of the data have a much larger weight than the architecture of the neural network.

Qi and Zhang (2005) state that there is substantial theoretical criticism towards neural networks and without adequate data pre-processing, its ability to predict even simple time series patterns of seasonality or trends is doubtful. Boyd and Kaastra (1996, p.6) define data pre-processing as the analysis and transformation of the input and output variables to minimise noise, emphasise important relationships, recognise trends, and "flatten" the distribution of the variable in order to assist the neural network in learning relevant patterns. The authors state that neural networks are "pattern matchers", and thus the representation of the data is essential in the design of the neural network. Data is rarely fed into the network in its raw form, and it is common for the data to be scaled between the upper and lower bounds of the transfer functions, which is usually between zero and one (Boyd & Kaastra, 1996).

The two most common data transformations as stated by Boyd and Kaastra (1996) are first differencing, and taking the natural logarithm of the variable. Linear trends can be removed by first differencing and data which can take on both small and large values can be smoothed using the logarithmic transformation (Boyd & Kaastra, 1996). Furthermore, logarithmic transformations may also be used to convert multiplicative or ratio relationships to relationships which are additive; this is said to shorten and improve network training (Masters, 1993).

Ratios of input variables are also another popular method to transform data. Not only do they have the ability to highlight important information, they also serve as a means to reduce degrees of freedom, as fewer input neurons are needed, thus reducing network training time (Boyd & Kaastra, 1996).

Logarithmic returns are used as the assumption that volatility cannot be negative is made which allows the activation functions to grasp the data more efficiently, enabling it to make legitimate inferences from the input series. Normalisation through the calculation of the logarithmic returns follows the following procedure:

$$R_i = \frac{P_{t+1}}{P_t} \tag{15}$$

where R_i is the daily return (normalised value), P_t is the current value of the SAVI at time *t* and P_t is the value of the SAVI at time *t+1*.

Boyd and Kaastra (1996) state that technical analysis can provide a neural network with a multitude of indicators, over and above first differencing, using ratios or taking the natural logarithm of input data. Some of these technical analysis tools include moving averages, oscillators, directional movement, as well as volatility filters. These tools have various functions, some of which include identifying overbought or oversold stock, identifying price trends, and indicating average stock prices over a period of time. Furthermore, the authors advise making use of a variety of indicators, as this reduces variable redundancy and also provides the network with the capacity to adapt to changing market conditions by means of periodic retraining.

3.4.6 Network Training, Testing and Validation Sets

This study makes use of the Levenberg Marquardt (LM) backpropagation algorithm for training the neural network. The algorithm provides a numerical solution to the problem of minimising a nonlinear function. This algorithm is one of the most popular tools for nonlinear minimum mean squares problems and due to its properties of quick convergence and stability, it has been used in many modelling problems (Hayami, Kuwahara, Matsumoto & Sakamoto, 2005). The LM algorithm exhibits the speed advantage of the Gauss-Newton algorithm (which is a modification of Newton's method of finding the minimum of a function, and used to solve nonlinear least squares problems) due to its ability to converge well even in cases where the "error surface is much more complex than the quadratic situation" (Wilamowski & Yu, 2011, p.2). The LM algorithm combines both the gradient descent method as well as the Gauss-Newton method for

neural network training. The gradient descent algorithm decreases along a function by taking steps in the opposite direction of the gradient of that function (Wiliamowski & Yu, 2011). Wilamowski and Yu (2011) state that even though the LM algorithm may be slower the Gauss-Newton algorithm, it still converges much more rapidly than the steepest decent method. A comparison of algorithms is made in Appendix C.

A neural network can be trained by presenting it with various input patterns which are contained in the data. The most widely used training algorithm is the backpropagation algorithm which may vary according to the network's architecture. A Backropagation network learns by example where the training phase is comprised of two passes. Connection weights are modified after the backward pass computes the error of the network based on the target outputs and passes these errors backward to the network. The neural network's performance is assessed by monitoring the convergence behaviour of the error. If the network "learns", the error will approach a minimum value (Saleemah, 2012).

Most neural networks require the time series to be divided into three distinct sets called the training, testing and validation sets. The network learns patterns in the data through the training set, which is also usually the largest segment of the entire data set. The testing set is used to evaluate the generalisation ability of the trained network (Boyd & Kaastra, 1996).

This study makes use of a time-delay neural network which is known to be used for time series analysis. In the case of where the type of neural network has not yet been selected, the network which performs best on the testing set should be selected. Furthermore, a final check is conducted on the performance of the trained network through the validation set.

Atiya et al. (1999) found a strong correlation between the training error and the testing error which suggests that there is good generalisation. The authors state that selecting an input set that gives a low training error will almost certainly result in a low testing error.

It is intuitive that larger neural networks require larger training datasets, and in turn should produce better forecasts and predictions. However, the concept of over fitting is more prominent when the model is large as stated by Bardina and Rajkumar (2003). The authors list a series of questions which should be addressed when selecting the training dataset for a given problem. These questions address factors like whether or not any transformations are needed for the

training dataset, the question of whether there are sufficient sample representations in all sub-classes, as well as questions relating to the number of layers needed in the neural network architecture. The debate of how many hidden neurons should be used, which has no hard and fast rule is also carefully looked into, together with the selection of an appropriate transfer function. Lastly, the length of time that training is required for the network is also listed by the authors as an important factor which needs to be taken into account when implementing a neural network.

This study uses the Tangens hyperbolicus (tanh) transfer function. The network's learning involves deviations from the average and thus the tanh function works best (Pissarenko, 2002). The tanh transfer function can be stated as:

$$transfer(x) = \tanh(x) = \frac{\sinh(x)}{\cosh(x)} = \frac{e^{x} - e^{-x}}{e^{x} + e^{-x}} \quad (16)$$

3.4.7 Number of Hidden Layers

The purpose of a hidden layer in a neural network is to equip the network with the ability to generalise. Theoretically, a neural network which only consists of one hidden layer with an adequate number of hidden neurons has the ability to approximate any continuous function. Increasing the number of hidden layers subsequently increases the computation time and also poses the danger of overfitting which leads to unsatisfactory forecasting performance (Boyd & Kaastra, 1996).

Boyd and Kaastra (1996) argue that the more weights relative to the size of the training set, the stronger the networks' ability to memorise any peculiarities of individual observations, which is actually detrimental to the model's use in actual forecasting, due to the validation set being lost. It is because of the aforementioned reasons that this study makes use of a single hidden layer, with ten hidden neurons, and two time delays. Later analysis proves that these specifications are the best for the time series in question.

3.4.8 Number of Hidden Neurons

The existing body of knowledge has not yet come to a concise conclusion as to the most efficient number of hidden neurons to use. Theory suggests that the final decision should be based on trial and error and through experimentation. With that being said however, there are some rules of thumb which have been articulated, which attempt to give the end user a reasonable amount of confidence in the selection of the number of hidden neurons.

Masters (1993) proposed the geometric pyramid rule whereby a network with three layers, comprised of n input neurons and m output neurons, would have a hidden layer of $\sqrt{n \times m}$ neurons. As stated previously, this study uses ten hidden neurons, and is justified later.

3.5. Experiment

In order to create enough data for the testing sample, a 70% versus 30% approach is used. As mentioned earlier, the Hyperbolic Tangent (tanh) transfer function is used for the hidden layer. This function has the advantage of an output interval of [-1, 1] and can thus be used in ANN's that are required to approximate functions that can take on negative values. A sigmoid transfer function restricts output in an interval of [0, 1].

Comparing output in the range [-1, 1] allows one to determine the next step and to determine the lowest RMSE. The basic design of the neural network consists of an input layer with the nodes equal to the number of inputs, one or multiple hidden layers and an output with a single output node.

3.5 Running the Multilayer Perceptron Programme

The NETLAB toolbox used[18] with MATLAB is used to construct the neural network. The MLP model is designed as follows:

1. Selecting the number of inputs
2. Select the number of hidden nodes
3. Selecting the number of delays
4. Select the activation function

The network is then trained by choosing the optimisation algorithm, entering the training and target data, choosing the number of epochs[19] for training and finally training the network. After the network is trained, it is then tested by using the training data set. Several network topologies are examined; networks with one and multiple hidden layers are considered, and the target network would be the one that exhibits the minimum error for both the training and testing patterns as is discussed in Seleemah (2012).

4. Empirical Results

The analysis of the results begins with the examination of the data, in particular, its return distribution and descriptive stats. Further analysis is done to determine whether or not the time series is stationary. The results of each forecasting tool is then presented and investigated, ending with a comparison of all the tools used.

As mentioned previously, in order to undertake the analysis that follows, the daily return series based upon the logarithmic return convention is generated. Furthermore, for estimation purposes, a total of 1687 observations are used, observations from 1 February 2007 to 31 October 2013 only being considered. The neural network however, makes use of the entire sample period as

[18] Appendix C provides the advanced code used in the implementation of the neural network in Matlab.

[19] An epoch is a step in the training process of a neural network.

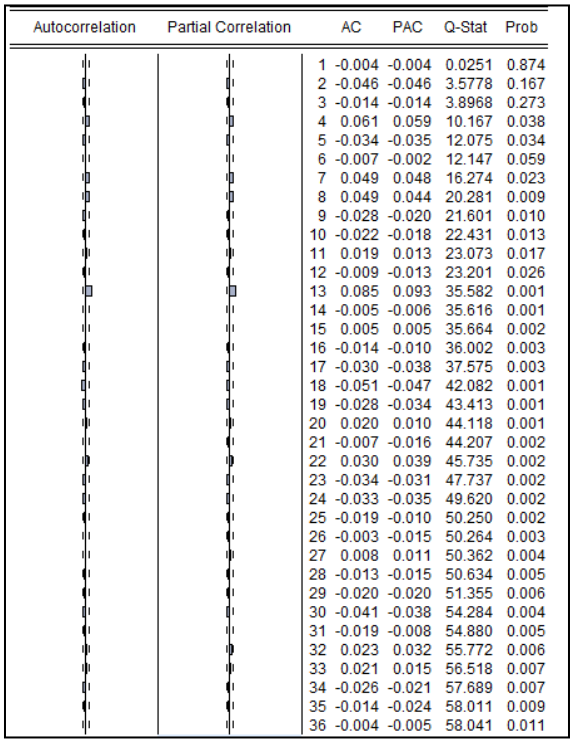
inputs to the neural network i.e. 1 February 2007 to 5 December 2013, with a total of 1713 observations.

4.1 Examining the Univariate Properties

If the daily returns on a stock index are independently and identically distributed (*iid*), then the assumption of returns being identically distributed ensures that the mean and variance of returns do not vary over time, which is consistent with a fixed probability assumption. The independence assumption ensures that speculative price changes are unrelated to each other at any point in time. It is therefore important to examine the distributional properties of returns, in order to ascertain whether or not the returns from the index are stationary.

Nagpaul (2005) states that stationarity is a critical assumption in time series models. Stationarity in the mean and variance implies that the likelihood of a specified loss will be the same for each day. A volatility index however, may show evidence of volatility clustering, and thus may not be *i.i.d.* Nonetheless, it is still useful to examine the validity of the assumptions.

Figure 8 below shows the correlogram of the return series, generated from the closing values of the SAVI, over the sample period from 2 Feb 2007 to 31 Oct 2013.



Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	-0.004	-0.004	0.0251	0.874
		2	-0.046	-0.046	3.5778	0.167
		3	-0.014	-0.014	3.8968	0.273
		4	0.061	0.059	10.167	0.038
		5	-0.034	-0.035	12.075	0.034
		6	-0.007	-0.002	12.147	0.059
		7	0.049	0.048	16.274	0.023
		8	0.049	0.044	20.281	0.009
		9	-0.028	-0.020	21.601	0.010
		10	-0.022	-0.018	22.431	0.013
		11	0.019	0.013	23.073	0.017
		12	-0.009	-0.013	23.201	0.026
		13	0.085	0.093	35.582	0.001
		14	-0.005	-0.006	35.616	0.001
		15	0.005	0.005	35.664	0.002
		16	-0.014	-0.010	36.002	0.003
		17	-0.030	-0.038	37.575	0.003
		18	-0.051	-0.047	42.082	0.001
		19	-0.028	-0.034	43.413	0.001
		20	0.020	0.010	44.118	0.001
		21	-0.007	-0.016	44.207	0.002
		22	0.030	0.039	45.735	0.002
		23	-0.034	-0.031	47.737	0.002
		24	-0.033	-0.035	49.620	0.002
		25	-0.019	-0.010	50.250	0.002
		26	-0.003	-0.015	50.264	0.003
		27	0.008	0.011	50.362	0.004
		28	-0.013	-0.015	50.634	0.005
		29	-0.020	-0.020	51.355	0.006
		30	-0.041	-0.038	54.284	0.004
		31	-0.019	-0.008	54.880	0.005
		32	0.023	0.032	55.772	0.006
		33	0.021	0.015	56.518	0.007
		34	-0.026	-0.021	57.689	0.007
		35	-0.014	-0.024	58.011	0.009
		36	-0.004	-0.005	58.041	0.011

Figure 8: Correlogram of the return series

It is evident from the above figure that autocorrelation could be adjudged to be present in the return series, as the Q-statistic is significant for the majority of the lag orders at a 95% confidence level. The autocorrelation coefficients drop to zero quickly, and thus do not appear to have an infinitely lived memory. Any shocks that occur will probably die out rapidly, rendering the series stationary. These results make economic sense for a volatility index, as the presence of autocorrelation suggests that the returns are governed by nonlinear processes which allow successive index changes to be linked through the variance.

4.2 Normality and Stable Moments

If the returns are normally distributed and exhibit stable moments, then the probability of a specified loss does not vary across days. In order to test for normality, it is useful to analyse the summary statistics of the return series, as well as examine a histogram of the return series, which will give an indication of where and how the data is spread, providing insight into whether or not the series is normal, before any formal tests are done. This section is also stated is order to check the consistency of the conclusions surrounding normality and stable moments of the return series.

Table 1 below displays the summary statistics for the return series. The mean of the daily return series is close to zero and the daily standard deviation is approximately 2.91%. The skewness measure is 0.510056, which is indicative of positive skewness, in contrast to the symmetric normal distribution. In addition, kurtosis is equal to 9.419747, which is much larger than a normal distribution, which yields a kurtosis value of 3. Hence, the distribution appears to be leptokurtic. Table 1 also displays the Jarque-Bera test statistic, which jointly examines the skewness and kurtosis to formulate a joint test statistic. The null hypothesis of normality is rejected at a 99% confidence level, as the *p-value* is less than 0.01.

Engle (2004) looked at daily levels of the S&P 500 composite index over a period of forty years. The authors found that the kurtosis over the entire period was dramatically high, which is indicative of extreme values being more substantial than that which would be expected if the return series followed a Normal distribution. The author found that the majority of information

stems from the previous day forecast, with the long run average variance proving to have a minute effect, in addition to new information producing a small change.

Figure 9 below suggests that the returns do appear to display some non-normality, in particular, fatter tails and more "peakedness" around the mean, both of which confirm the results from the summary statistics.

Table 1: Summary statistics for the return series

Mean	-0.000166
Median	-0.000479
Maximum	0.194477
Minimum	-0.243413
Std. Dev.	0.029059
Skewness	0.510056
Kurtosis	9.419747
Jarque-Bera	2970.089
Probability	0.000000

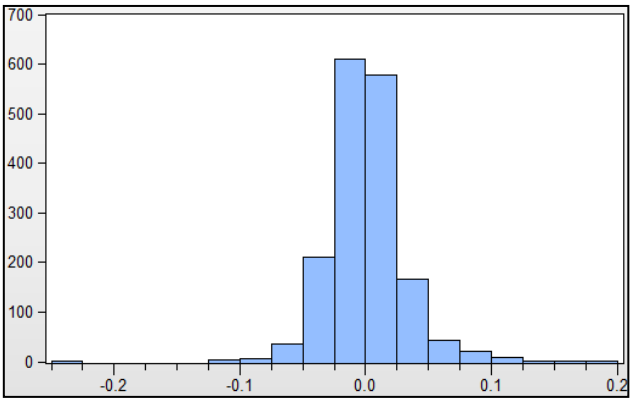


Figure 9: Histogram of the return series

Further analysis may be done by examining the non-stationary nature of the first two statistical moments. The leptokurtosis described above could be a result of the time varying nature of these moments.

Figure 10 below depicts a plot of the daily return series; which shows a fairly constant mean across sub samples of the data, which is consistent with other studies of financial returns, where the mean is often found to be stationary. However, the figure also suggests that there are periods of increased volatility, followed by more tranquil periods, which is consistent with the notion of volatility clustering.

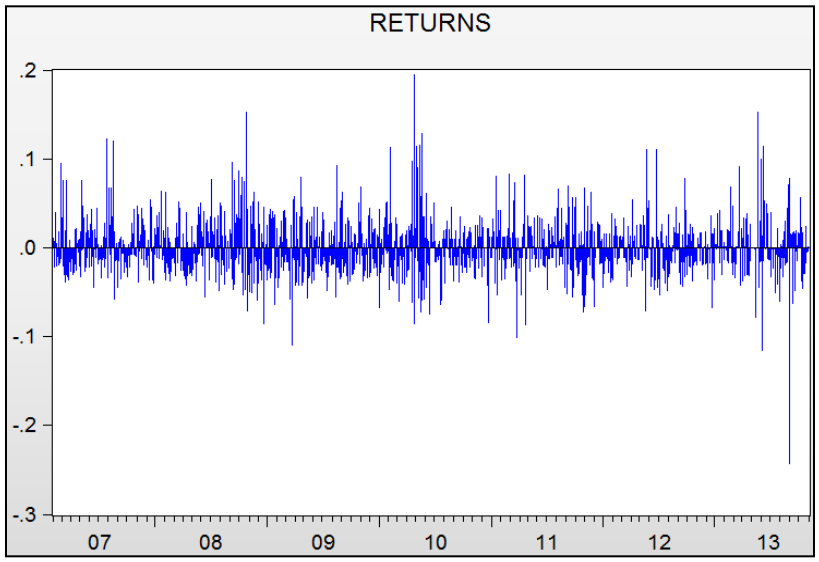


Figure 10: A plot of the daily return series

4.3 Stationarity and Unit Root tests

Nelson and Plosser (1982) claim that the order of integration[20] of a particular variable is a simple characteristic of all variables in the context of business cycles[21]. The authors presented statistical

[20] The order of integration is the amount of times a variable needs to be differenced until it becomes covariance stationary.

[21] It can be argued that the SAVI changes according to business cycles, as does investor sentiment, in order to adapt according to the new business environment.

proof of the existence of a stochastic trend in about 80% of the aggregate variables of the US economy. Furthermore, the authors made use of the Dickey and Fuller (1979) procedure and the correlogram. This study makes use of the Augmented Dickey-Fuller test to assess the stationarity of the return series in question.

The Augmented Dickey-Fuller test constructs a parametric correction for higher order correlation by assuming that the series follows an AR (p) process and adding lagged difference terms of the dependent variable to the right hand side of the test regression, resulting in the following equation:

$$\Delta y_t = \alpha y_{t-1} + \delta x_t + \beta_1 \Delta y_{t-1} + \beta_2 \Delta y_{t-2} + \cdots + \beta_p \Delta y_{t-p} + \varepsilon_t \quad (17)$$

This study makes use of the Augmented Dickey-Fuller test with a constant and a linear trend. The results presented in Table 2 below suggest that one would reject the null hypothesis that the return series has a unit root at a 99% confidence level, as the ADF test statistic is equal to -41.20505 and the *p-value* is essentially zero. Thus, the results suggest that the return series is stationary.

Further confirmation of the above results arises from the Phillips-Perron (PP) test, using the Bartlett kernel for the Spectral estimation method. The results are presented in Table 3 below and suggest that one would reject the null hypothesis of the return series having a unit root at the 99% confidence level.

Table 2: Augmented Dickey Fuller test statistics

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-41.20505	0.0000
Test critical values:	1% level	-3.963495	
	5% level	-3.412477	
	10% level	-3.128189	

*MacKinnon (1996) one-sided p-values.

Augmented Dickey-Fuller Test Equation
Dependent Variable: D(RETURNS)
Method: Least Squares
Date: 12/09/13 Time: 11:49
Sample (adjusted): 2/05/2007 10/31/2013
Included observations: 1686 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
RETURNS(-1)	-1.004361	0.024375	-41.20505	0.0000
C	0.000953	0.001418	0.672218	0.5015
@TREND(2/01/2007)	-1.33E-06	1.46E-06	-0.916646	0.3595

Table 3: Phillips-Perron test statistics

		Adj. t-Stat	Prob.*
Phillips-Perron test statistic		-41.21088	0.0000
Test critical values:	1% level	-3.963495	
	5% level	-3.412477	
	10% level	-3.128189	

*MacKinnon (1996) one-sided p-values.

Residual variance (no correction)	0.000844
HAC corrected variance (Bartlett kernel)	0.000821

Phillips-Perron Test Equation
Dependent Variable: D(RETURNS)
Method: Least Squares
Date: 12/09/13 Time: 11:58
Sample (adjusted): 2/05/2007 10/31/2013
Included observations: 1686 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
RETURNS(-1)	-1.004361	0.024375	-41.20505	0.0000
C	0.000953	0.001418	0.672218	0.5015
@TREND(2/01/2007)	-1.33E-06	1.46E-06	-0.916646	0.3595

4.4 Determining the ARMA structure of the Data Generating Process

It is necessary to establish the autoregressive moving average structure of the data generating process (DGP). The general ARMA (p, q) process is defined as follows:

$$y_t = \delta + \sum_{i=1}^{p} \emptyset_i y_{t-i} + \varepsilon_t + \sum_{j=1}^{q} \theta_j \varepsilon_{t-j} \quad (18)$$

A lag order of 4 is used for the maximum AR term as well as the maximum MA term. It is important to note that no seasonal autoregressive (SAR) or seasonal moving average (SMA) terms are used as daily data is used; Box and Jenkins (1970) recommend the use of SAR and SMA terms for monthly or quarterly data with systematic seasonal movements.

The lag order chosen for the ARMA structure is based upon minimising the model selection criterion. This study bases the ARMA structure on minimising the Akaike Information Criterion (AIC). In this case, for the AR component, p is equal to 4 and the for the MA component, q is equal to 4. Therefore, the preferred equation specification for the data generating process should be given by:

$$y_t = \delta + \emptyset_1 y_{t-1} + \emptyset_2 y_{t-2} + \emptyset_3 y_{t-3} + \emptyset_4 y_{t-4} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \theta_3 \varepsilon_{t-3} + \theta_4 \varepsilon_{t-4} \quad (19)$$

Table 4 below outlines the corresponding AIC values for the various lag order combinations. It is evident from Table 4 that the minimum AIC value is obtained when both the autoregressive and moving average component is 4.

Table 5 below depicts the regression results for the ARMA (4, 4) model of the data generating process. The table displays the coefficients attaching to the corresponding autoregressive and moving average terms. It can be seen that all of these terms are statistically significantly different from zero at the 95% confidence level.

Table 4: AIC Values for various lag order combinations

AR / MA	0.000000	1.000000	2.000000	3.000000	4.000000
0.000000	-4.237038	-4.235865	-4.236511	-4.235575	-4.237923
1.000000	-4.235865	-4.235011	-4.235363	-4.236487	-4.237141
2.000000	-4.236767	-4.235614	-4.238669	-4.240393	-4.239232
3.000000	-4.235779	-4.236472	-4.240415	-4.239228	-4.238584
4.000000	-4.237490	-4.237148	-4.238666	-4.237896	-4.240912

Table 5: Estimation output - ARMA (4, 4) model

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.000142	0.000720	-0.197084	0.8438
AR(1)	1.022849	0.060868	16.80447	0.0000
AR(2)	-0.458445	0.049485	-9.264386	0.0000
AR(3)	1.012088	0.046195	21.90894	0.0000
AR(4)	-0.851575	0.057477	-14.81585	0.0000
MA(1)	-1.020199	0.055198	-18.48252	0.0000
MA(2)	0.417491	0.047421	8.803852	0.0000
MA(3)	-0.992832	0.043856	-22.63845	0.0000
MA(4)	0.876093	0.051263	17.09007	0.0000

R-squared	0.013848	Mean dependent var	-0.000165
Adjusted R-squared	0.009135	S.D. dependent var	0.029087
S.E. of regression	0.028954	Akaike info criterion	-4.240912
Sum squared resid	1.403331	Schwarz criterion	-4.211884
Log likelihood	3577.728	Hannan-Quinn criter.	-4.230161
F-statistic	2.938280	Durbin-Watson stat	2.015842
Prob(F-statistic)	0.002904		

Inverted AR Roots	.90+.30i	.90-.30i	-.39-.89i	-.39+.89i
Inverted MA Roots	.91-.31i	.91+.31i	-.40-.89i	-.40+.89i

4.4.1 ARMA Equation Diagnostics

After estimating the ARMA (4, 4) model for the data generating process, it is necessary to examine the ARMA equation diagnostics. This study looks at the roots, correlogram and impulse response as an analysis of the ARMA model.

Figure 11 below displays the inverse roots of the AR and MA characteristic polynomial, where the roots are plotted in a complex plane, with the horizontal axis measuring the real part, and the vertical axis reflecting the imaginary part of each root. The figure depicts that the ARMA process is stationary as all the AR roots lie inside of the unit circle. Furthermore, the fact that all the MA roots lie inside the unit circle renders the ARMA process invertible as well. The same conclusions can be made when examining Table 6, which displays all roots in order of decreasing modulus, indicating that no root lies outside the unit root circle, and that the ARMA model is invertible.

From both the graphical and tabular views of the inverse roots, one can conclude that the estimated ARMA model of the data generating process presented above is in fact (covariance) stationary and invertible.

Figure 12 displays a graphical view of the "actual" and ARMA model correlogram. The figure compares the autocorrelation pattern of the structural residuals and those of the estimated model, for a certain number of periods. It is evident from the figure that even though the ARMA model in question is stationary and invertible, it does appear that the ARMA model could be improved as the residual and estimated autocorrelations and partial autocorrelations are not always close together.

Table 7 represents a tabular view of the actual and ARMA model correlogram, resulting in the same conclusion drawn as when analysing Figure 12, there is evidence of a degree of misspecification in relation to the estimated ARMA model of the data generating process.

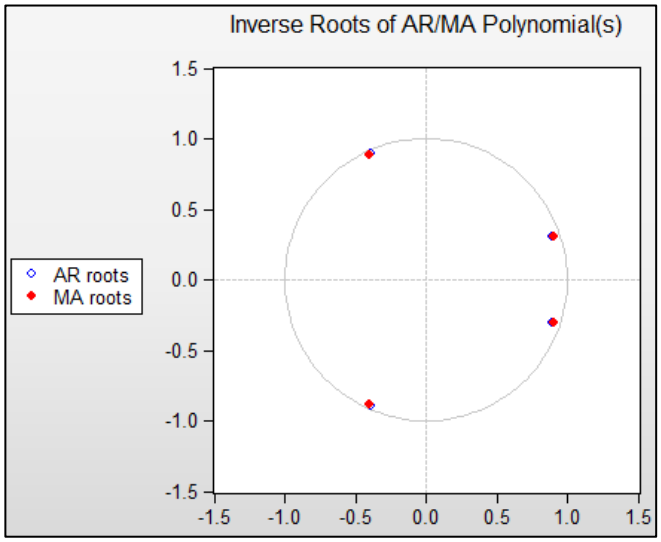


Figure 11: Graphical view of the inverse roots of the AR and MA characteristic polynomial

Table 6: Tabular view of the inverse roots of the AR and MA characteristic polynomial

AR Root(s)	Modulus
-0.386449 ± 0.894562i	0.974466
0.897873 ± 0.301017i	0.946989
No root lies outside the unit circle. ARMA model is stationary.	
MA Root(s)	Modulus
-0.399727 ± 0.889250i	0.974960
0.909826 ± 0.306411i	0.960037
No root lies outside the unit circle. ARMA model is invertible.	

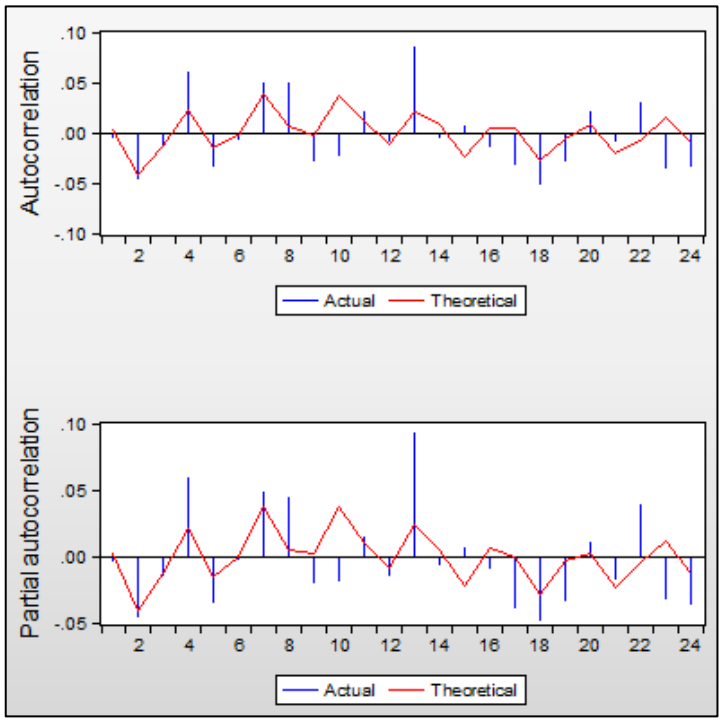


Figure 12: Graphical view of the actual and ARMA model correlogram

Table 7: Tabular view of the actual and ARMA model correlogram

	Autocorrelation				Partial Autocorrelation		
	Actual	Model	Difference		Actual	Model	Difference
0	1.000	1.000	0.000	0	1.000	1.000	0.000
1	-0.004	0.002	-0.006	1	-0.004	0.002	-0.006
2	-0.045	-0.041	-0.004	2	-0.045	-0.041	-0.004
3	-0.014	-0.013	-0.001	3	-0.014	-0.013	-0.001
4	0.061	0.023	0.038	4	0.059	0.021	0.037
5	-0.034	-0.014	-0.019	5	-0.035	-0.016	-0.019
6	-0.007	-0.003	-0.004	6	-0.002	-0.001	-0.001
7	0.049	0.038	0.012	7	0.048	0.037	0.011
8	0.049	0.006	0.043	8	0.045	0.005	0.040
9	-0.028	-0.002	-0.026	9	-0.020	0.002	-0.021
10	-0.022	0.036	-0.058	10	-0.018	0.037	-0.055
11	0.020	0.011	0.009	11	0.014	0.010	0.004
12	-0.009	-0.012	0.003	12	-0.014	-0.008	-0.005
13	0.085	0.021	0.064	13	0.093	0.023	0.070
14	-0.005	0.008	-0.013	14	-0.006	0.004	-0.010
15	0.007	-0.023	0.030	15	0.006	-0.022	0.028
16	-0.014	0.004	-0.017	16	-0.009	0.006	-0.015
17	-0.031	0.005	-0.035	17	-0.038	-0.001	-0.037
18	-0.052	-0.027	-0.024	18	-0.048	-0.029	-0.019
19	-0.027	-0.006	-0.021	19	-0.033	-0.004	-0.029
20	0.020	0.008	0.012	20	0.010	0.002	0.008
21	-0.008	-0.021	0.013	21	-0.017	-0.024	0.008
22	0.030	-0.008	0.038	22	0.039	-0.004	0.043
23	-0.034	0.014	-0.049	23	-0.031	0.011	-0.042
24	-0.034	-0.009	-0.024	24	-0.035	-0.012	-0.023

Further analysis is done by examining the impulse response function, which traces the response of the ARMA part of the estimated equation to shocks in the innovation term. Effectively, the impulse response function traces the response to an isolated shock in the innovation term. The accumulated response, as the name suggests, is the accumulated sum of the impulse responses. This can also be interpreted as the response to a step impulse, whereby a single identical shock occurs in ever period, emanating from the first period. This study uses the default value for the number of periods as 24, and defines the shock as being two standard deviations, utilising the standard error of the regression for the estimated equation.

Figure 13 displays the impulse response function of the ARMA model, with Table 8 providing a tabular view. It is evident from both the graphical and tabular views that the ARMA model is stationary, as the impulse responses asymptote to zero, whilst the accumulated values asymptote to their long run value of 2.039770. The asymptotic values are represented by dotted lines in figure 13.

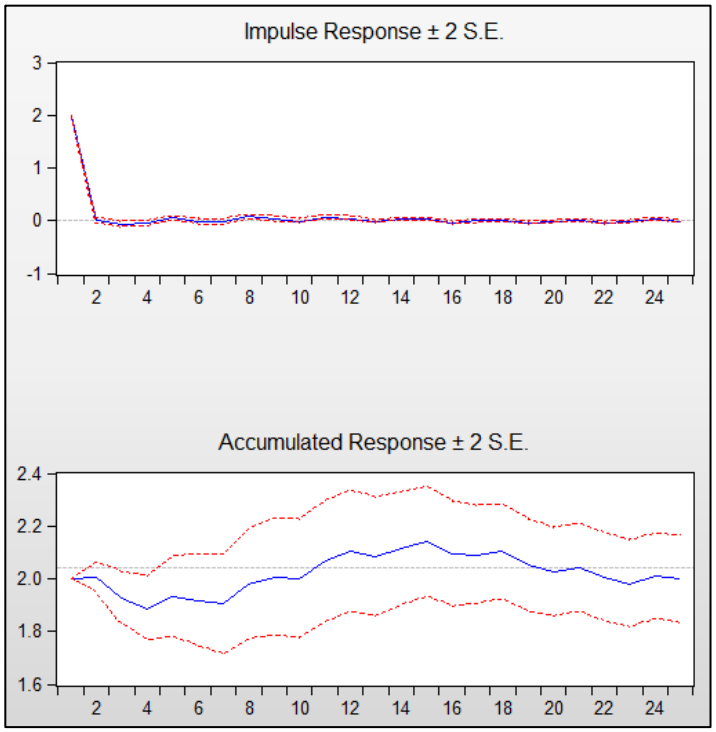


Figure 13: Impulse response function of ARMA model

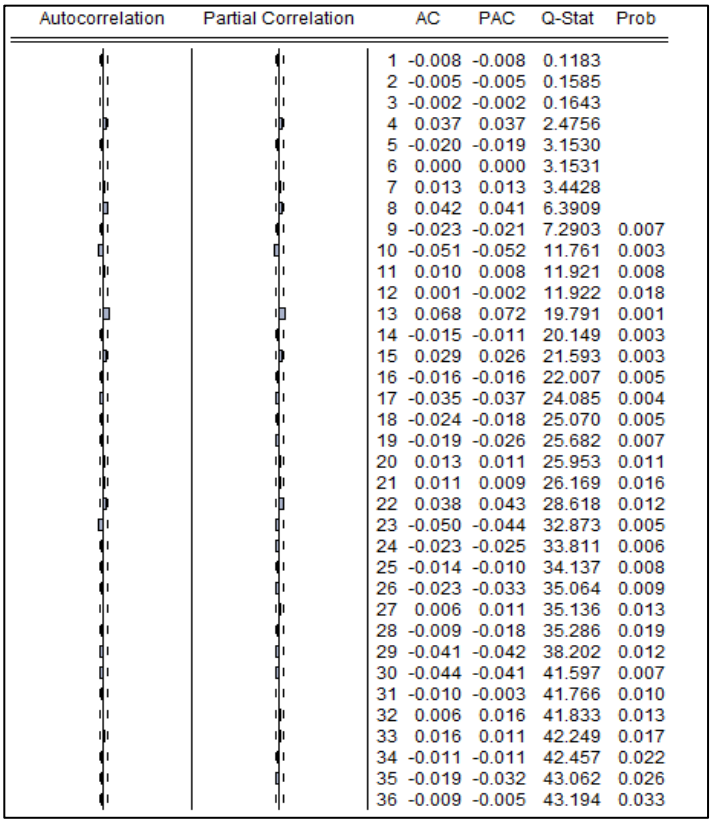
Table 9 below displays the residuals from the model. It can be noted that all of the Q-statistics are significant at the 95% confidence level. This suggests that the ARMA model can be

improved, as there is still serial correlation left in the residuals. The current ARMA structure is selected based on the minimum AIC value. Using other ARMA structures yielded spurious and insignificant autoregressive and moving average terms. Furthermore, by analysing the correlogram, one can infer that there are no significant "surprises" which may cause any drastic alteration of the results. This study does not examine other methods of selecting the ARMA model, but emphasis should be placed on minimising the AIC.

Table 8: Tabular view of ARMA impulse response function

Period	Response	Std.Err.	Accumulated	Std.Err.
1	2.000000	(0.00000)	2.000000	(0.00000)
2	0.005301	(0.02989)	2.005301	(0.02989)
3	-0.076485	(0.02886)	1.928815	(0.04789)
4	-0.042151	(0.02856)	1.886664	(0.06120)
5	0.046349	(0.02566)	1.933013	(0.07520)
6	-0.015191	(0.02551)	1.917822	(0.08730)
7	-0.014315	(0.02390)	1.903507	(0.09595)
8	0.075127	(0.02385)	1.978634	(0.10410)
9	0.028561	(0.02306)	2.007195	(0.11089)
10	-0.006779	(0.01886)	2.000416	(0.11339)
11	0.068198	(0.01865)	2.068614	(0.11475)
12	0.037794	(0.01843)	2.106408	(0.11542)
13	-0.023790	(0.01560)	2.082617	(0.11183)
14	0.033134	(0.01557)	2.115752	(0.10728)
15	0.024973	(0.01546)	2.140725	(0.10419)
16	-0.045909	(0.01491)	2.094817	(0.09883)
17	-0.004612	(0.01494)	2.090205	(0.09318)
18	0.013388	(0.01392)	2.103592	(0.09063)
19	-0.051922	(0.01385)	2.051670	(0.08779)
20	-0.024819	(0.01384)	2.026851	(0.08447)
21	0.015894	(0.01281)	2.042745	(0.08378)
22	-0.036315	(0.01274)	2.006430	(0.08339)
23	-0.025335	(0.01272)	1.981095	(0.08179)
24	0.027956	(0.01238)	2.009052	(0.08156)
25	-0.010079	(0.01239)	1.998972	(0.08200)
LR	0.000000	(0.00000)	2.039770	(0.05962)

Table 9: Correlogram of the residuals - ARMA model



Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	-0.008	-0.008	0.1183	
		2	-0.005	-0.005	0.1585	
		3	-0.002	-0.002	0.1643	
		4	0.037	0.037	2.4756	
		5	-0.020	-0.019	3.1530	
		6	0.000	0.000	3.1531	
		7	0.013	0.013	3.4428	
		8	0.042	0.041	6.3909	
		9	-0.023	-0.021	7.2903	0.007
		10	-0.051	-0.052	11.761	0.003
		11	0.010	0.008	11.921	0.008
		12	0.001	-0.002	11.922	0.018
		13	0.068	0.072	19.791	0.001
		14	-0.015	-0.011	20.149	0.003
		15	0.029	0.026	21.593	0.003
		16	-0.016	-0.016	22.007	0.005
		17	-0.035	-0.037	24.085	0.004
		18	-0.024	-0.018	25.070	0.005
		19	-0.019	-0.026	25.682	0.007
		20	0.013	0.011	25.953	0.011
		21	0.011	0.009	26.169	0.016
		22	0.038	0.043	28.618	0.012
		23	-0.050	-0.044	32.873	0.005
		24	-0.023	-0.025	33.811	0.006
		25	-0.014	-0.010	34.137	0.008
		26	-0.023	-0.033	35.064	0.009
		27	0.006	0.011	35.136	0.013
		28	-0.009	-0.018	35.286	0.019
		29	-0.041	-0.042	38.202	0.012
		30	-0.044	-0.041	41.597	0.007
		31	-0.010	-0.003	41.766	0.010
		32	0.006	0.016	41.833	0.013
		33	0.016	0.011	42.249	0.017
		34	-0.011	-0.011	42.457	0.022
		35	-0.019	-0.032	43.062	0.026
		36	-0.009	-0.005	43.194	0.033

4.5 Testing for ARCH Effects

ARCH and GARCH models have been discussed extensively in Chapter 2. The literature suggests that both types of models are designed to eliminate the systematically changing variance from the data which is responsible for most of the leptokurtosis identified previously.

Equation (19) above describes the appropriate specification for the data generating process which underlie the daily returns on the SAVI, where the current period's returns are established by an ARMA (4, 4) process.

The ARCH foundation is based on modelling the square of the residual return. ARCH effects are tested by examining whether or not the volatility of the daily returns on the SAVI and the squares of the daily returns are serially correlated. This is done by running a regression of

squared Ordinary Least Squares (OLS) residuals which originate from Equation (19), on the lagged squared residuals. The following null hypothesis of "no ARCH effects" is used:

$$H_{0:}\ \alpha_1=\alpha_2=\cdots=\alpha_q=0 \tag{20}$$

Table 10 below displays the analysis for ARCH (1) effects. The Lagrange Multiplier (LM) test yields a test statistic of 22.11927, with a *p-value* of 0. The null hypothesis that there are no ARCH effects in Equation (19) is thus rejected at a 95% level of significance. The F-Statistic yields a similar conclusion.

It is tempting at this point to assume that by rejecting the null hypothesis that there are no ARCH effects implies that the conditional variance of Equation (19) is non constant. Further analysis through the Breusch Godfrey serial correlation (LM) test however suggests that the null hypothesis that the residuals are serially uncorrelated cannot be rejected at a 95% confidence level as the *p-value* is greater than 0.05. Table 10 below displays this result.

The above mentioned results and diagnostic checks may appear to be trivial at first; however, this analysis is imperative in order to understand the inputs and procedures used by the various forecasting techniques. Furthermore, by scrutinising the data in such a manner, one is able to make more intelligent inferences about the forecasting results obtained. In addition, the forecasting tools used may be adapted based on the analysis of the data, more suitable forecasting methods may be selected based on the data and its characteristics, and avenues for further research may also be exploited which may stem from such an analysis .

Table 10: Breusch-Godfrey serial correlation LM test

Breusch-Godfrey Serial Correlation LM Test:

F-statistic	0.187005	Prob. F(1,1673)	0.6655
Obs*R-squared	0.187858	Prob. Chi-Square(1)	0.6647

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	6.38E-07	0.000720	0.000886	0.9993
AR(1)	0.009768	0.064948	0.150394	0.8805
AR(2)	-0.000669	0.049507	-0.013513	0.9892
AR(3)	-0.002638	0.046595	-0.056617	0.9549
AR(4)	-0.004694	0.058507	-0.080223	0.9361
MA(1)	-0.004793	0.056321	-0.085099	0.9322
MA(2)	-0.002847	0.047870	-0.059479	0.9526
MA(3)	0.004079	0.044857	0.090943	0.9275
MA(4)	0.002579	0.051625	0.049958	0.9602
RESID(-1)	-0.013343	0.030855	-0.432440	0.6655

R-squared	0.000112	Mean dependent var	1.10E-05
Adjusted R-squared	-0.005267	S.D. dependent var	0.028885
S.E. of regression	0.028961	Akaike info criterion	-4.239836
Sum squared resid	1.403174	Schwarz criterion	-4.207582
Log likelihood	3577.822	Hannan-Quinn criter.	-4.227889
F-statistic	0.020751	Durbin-Watson stat	1.999303
Prob(F-statistic)	1.000000		

4.6 Historical Average Model (HAM)

The historical average model is computed using equation (4). As discussed earlier, this forecasting technique is considered a naive approach to forecasting. It is the simplest method used in this study, and results are stated merely as a benchmark to compare the other models to. Ladokhin (2009) reported a RMSE of 0.1235, which was the worst performing model. This makes intuitive sense due its simplicity, and mere disregard for other factors which could affect volatility. The empirical analysis done in this study however, yields a RMSE of 0.0253. This result should be interpreted with caution however, as the sample size in this study is fairly small, due to SAVI itself being fairly new. A longer time period, with a larger sample size would most likely alter the forecastability of not only the HAM, but all the other forecasting tools as well. In addition, a longer time period would also allow for more random shocks to appear and evolve over time which would be much more difficult to predict by using the simple historical average model. This would in turn, leave the more sophisticated forecasting tools with a higher probability of being able to forecast the volatility in question, due to their ability to handle and manipulate more complex data, as discussed previously.

4.7 Exponential Smoothing

The parameters of the ES model are obtained by minimising the sum of squared errors.[22] Although there are various forms of estimating an ES model, this study uses the "Single Smoothing" method, as it is more common to use for a random series which exhibits no trend or seasonal patterns. The average of the first $\left(\frac{(T+1)}{2}\right)$ observations of the series is used to start the recursion, where *T* is equal to the number of observations in the sample.

Table 11 below displays the forecasting power of the ES method. The estimated parameter for α, of 0.001 is not close to 1, and hints at the fact that perhaps the return series being forecasted is not by definition, a random walk, whereby the current value is the best predictor of future values.

[22] This is an automated function in EViews, whereby the parameters are constrained to be between zero and one. Bowerman and O'Connell (1979) recommend using values from 0.01 to 0.3 for a one parameter model.

Furthermore, the RMSE of 0.029063 indicates the model's forecasting ability, suggesting that the model can be improved, as a value below 0.01 indicates a reasonably accurate model. Nonetheless, the forecasting ability based on this measure is more promising than the results presented by Ladokhin (2009), who presented a RMSE of 0.0886.

Table 11: Single Exponential Smoothing results

Parameters: Alpha		0.0010
Sum of Squared Residuals		1.424940
Root Mean Squared Error		0.029063
End of Period Levels:	Mean	-0.000295

4.8 Exponentially Weighted Moving Average

As discussed earlier, this study uses the methodology presented by Shumway and Stoffer (2011) in order to forecast using EWMA, by adapting the ES model using the Holt-Winters method, setting the parameter for α equal to $1-\lambda$, and the parameters β and Φ equal to zero. This study follows RiskMetrics™ and uses a λ of 0.94, this making $\alpha=0.06$

The Holt-Winters method is a two-parameter method which restricts any seasonal factors from varying over time, such that any positive seasonal factors remain in the forecasts.

Table 12 below displays the results from forecasting using EWMA. The value of α is computed as 0.06, rather than being estimated based on minimising the sum of squared errors. The RMSE is 0.029529, which is almost identical to the ES method, but indicates a slightly worse forecasting model. Ladokhin (2009) reported a RMSE of 0.0784 which indicates little or no forecasting ability.

Table 12: EWMA results

Parameters: Alpha		0.0600
Beta		0.0000
Sum of Squared Residuals		1.470998
Root Mean Squared Error		0.029529
End of Period Levels:	Mean	-0.003458
	Trend	2.87E-05

4.10 ARMA Model

The data generating process was developed and estimated earlier. The ARMA (4, 4) model is thus used as a forecasting tool. Brooks (2008) recommends defining the problem as a nonlinear least squares (NLS) problem. This study adopts a similar approach, whereby the output is interpreted in the same manner as when the data generating process was defined in Table 5 above. Brooks (2008) states that the ARMA model is not based on any economic or financial theory and thus interpreting the individual parameter estimates is unrewarding. The author recommends examining the plausibility of the model as whole, in order to ascertain whether or not it describes the data well and manufactures accurate forecasts.

Table 13 below displays the forecasting ability of the ARMA (4, 4) model, relative to the data used in this study. The RMSE is 0.028876 which is slightly better than the forecastability of the EWMA model and the ES model. It is tempting at this point to assume forecastability of the SAVI using the ARMA (4, 4) model, however, at closer inspection, the Thiel Innequality Coefficient (Thiel's U)[23] suggests otherwise. The value of 0.894673, which can be argued as being fairly close to 1, suggests that forecasting using the ARMA model is only slightly better than a naive guess. Ladokhin (2009) reported a RMSE of 0.0890 which suggests poor forecastability.

Table 13: ARMA results

Forecast: RETURNSARMA	
Actual: RETURNS	
Forecast sample: 2/01/2007 10/31/2013	
Adjusted sample: 2/08/2007 10/31/2013	
Included observations: 1683	
Root Mean Squared Error	0.028876
Mean Absolute Error	0.020337
Mean Abs. Percent Error	123.8046
Theil Inequality Coefficient	0.894673
Bias Proportion	0.000000
Variance Proportion	0.803623
Covariance Proportion	0.196377

[23] Thiel's U provides a means to assess how well a time series of estimated values compares to a corresponding time series of observed values (Thiel, 1966)

4.11 ARCH Model

The necessary diagnostic checks before making use of an ARCH model are presented above, and not restated here. Once these diagnostic checks are done and the data generating process is defined, the implementation of an ARCH model is fairly simple.

Table 14 below presents the results. The RMSE is 0.029018 which shows inferior forecastability than the ARMA model, but slightly better forecastability than the EWMA model. The forecastability of the ARCH model presented is similar to the forecastability of the ES model. Ladokhin (2009) reported a RMSE of 0.1029 for a simple ARCH model, indicating poor forecastability.

Table 14: ARCH results

Forecast: RETURNSARCH	
Actual: RETURNS	
Forecast sample: 2/01/2007 10/31/2013	
Adjusted sample: 2/08/2007 10/31/2013	
Included observations: 1683	
Root Mean Squared Error	0.029018
Mean Absolute Error	0.020320
Mean Abs. Percent Error	118.8914
Theil Inequality Coefficient	0.904483
Bias Proportion	0.000803
Variance Proportion	0.817815
Covariance Proportion	0.181382

Based on the analysis from Appendix A, the daily long run variance using the ARCH model can be computed as follow:

$$V_{L=\frac{\hat{\alpha}_0}{1-\hat{\alpha}_0-\hat{\beta}_1}} \tag{21}$$

$$= \frac{0.000642}{1 - 0.283889}$$

$$= 0.000897$$

where $\hat{\beta}_1$ is zero and specified in a GARCH model. This implies that an estimate of the daily long run volatility of the SAVI would be equal to 0.029949, which is fairly similar to the standard deviation of the entire sample period of 0.028956.

4.12 GARCH

As with the ARCH model presented above, all pre-diagnostic checks are presented above, which are the same for the ARCH model.

Table 15 displays the forecasting results of the GARCH (1,1) model used. The RMSE is 0.028973 which is only slightly better than the forecastability of the ARCH model presented above. Thiel's U is 0.873324 which is close to 1, indicating poor forecastability of the model, but displaying slightly improved forecastability compared to the ARCH model. Ladhokin (2009) presented a RMSE of 0.1011 and 0.1014 for the two GARCH models used. In addition, the RMSE for the ARCH and GARCH model used are similar, as is the case in this study.

Table 15: GARCH (1, 1) results

Forecast: RETURNSGARCH	
Actual: RETURNS	
Forecast sample: 2/01/2007 10/31/2013	
Adjusted sample: 2/08/2007 10/31/2013	
Included observations: 1683	
Root Mean Squared Error	0.028973
Mean Absolute Error	0.020338
Mean Abs. Percent Error	133.3911
Theil Inequality Coefficient	0.873324
Bias Proportion	0.001353
Variance Proportion	0.754727
Covariance Proportion	0.243921

Following the same approach as the analysis done for the ARCH model, an analysis of Appendix B suggests the following daily long run variance using the GRACH model, which can be computed as follows: 0.000105, $\hat{\alpha}_1 = 0.156567$ and $\hat{\beta}_1 = 0.725849$.

$$V_{L=\frac{\hat{\alpha}_0}{1-\hat{\alpha}_0-\hat{\beta}_1}}$$

$$= \frac{0.000105}{1 - 0.156567 - 0.725849}$$

$$= 0.000893$$

This implies that an estimate of the daily long run volatility of the SAVI would be equal to 0.029883, which is fairly similar to the standard deviation of the entire sample period of 0.028956.

4.13 TDNN

As discussed previously, this study uses the Levenberg-Marquardt algorithm for training the neural network. Furthermore, the data set is split into three sets as follows:

1. 60% is used for training.
2. 20% is used to validate that the network is generalising, as well as to stop training the network before overfitting.
3. The last 20% is used as a completely independent test of network generalisation.

The use of a TDNN in this study is defined as a Nonlinear Autoregressive (NAR) time series problem, whereby a series *Y(t)* is predicted, given *d* past values of *Y(t).*

Figure 14 below depicts the TDNN used in this study. The network is generated and trained in open loop form as depicted in Figure 14; this allows the network to be supplied with correct feedback inputs, in order to generate the correct feedback outputs. After the network is trained, it may be converted to closed loop form if this is required by the application.

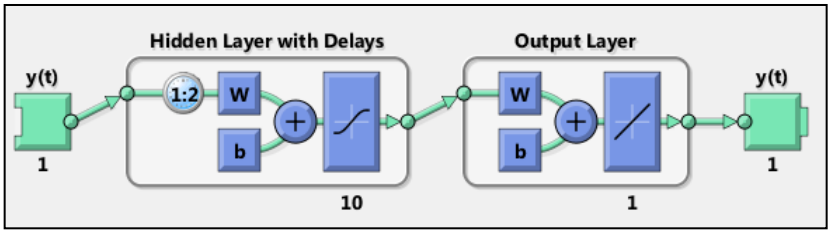


Figure 14 : Graphical representation of TDNN

Figure 15 below depicts the algorithm and progress of the TDNN in question. The network stopped at iteration 11. The training which stopped at iteration 11 could be due to the network using the Levenberg-Marquardt (LM) backpropagation algorithm, as discussed earlier.

Figure 16 displays the performance of the training, with a plot of the training errors, validation errors, and test errors. The final mean squared error is small, validating the results. Furthermore, the figure depicts that the test set error and the validation set error have similar characteristics. Lastly, the best validation performance occurred by iteration 5, whereby no significant overfitting seems to have occurred, another advantage of using the Levenberg-Marquardt (LM) backpropagation algorithm.

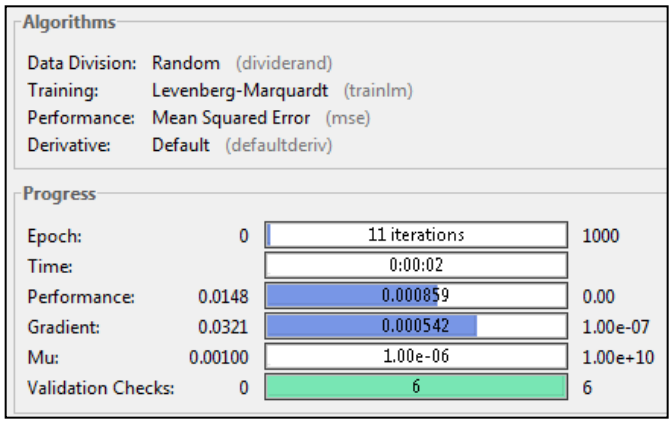


Figure 15: Algorithm and progress of TDNN

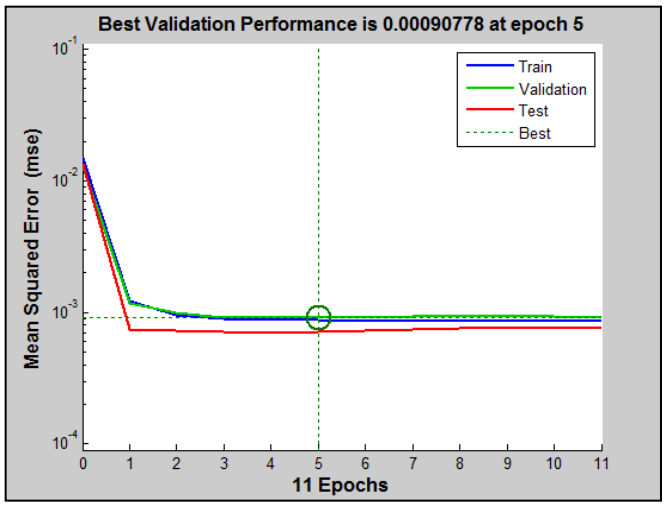


Figure 16: Best validation performance

Figure 17 below shows a plot of the error autocorrelation function. The figure illustrates the relation between the prediction errors through time. A perfect prediction model would consist of one lag which is greater than zero, which occurs at zero lag, which is in fact, the mean squared error. This result would suggest that the prediction errors are completely uncorrelated with each other and are merely white noise. Figure 17 shows that there is only slight correlation between the prediction errors, with the correlations falling within the 95% confidence limit around zero. This suggests that the model is fairly adequate.

Figure 18 below represents the inputs, targets, and errors against time. In addition, the figure also indicates which time points were selected for training, testing, and validation.

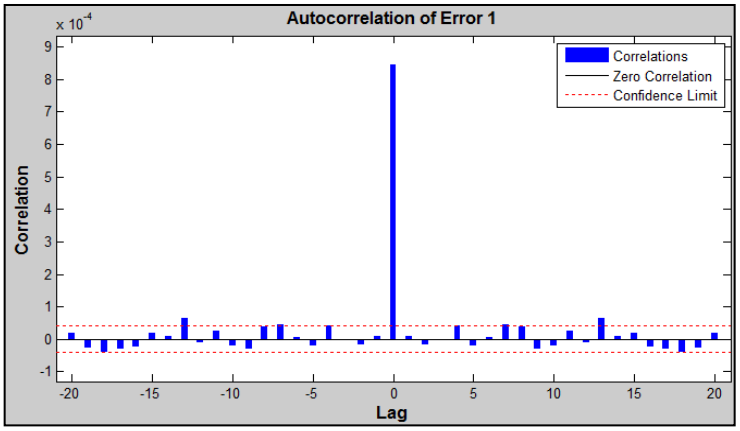


Figure 17: Plot of the error autocorrelation function

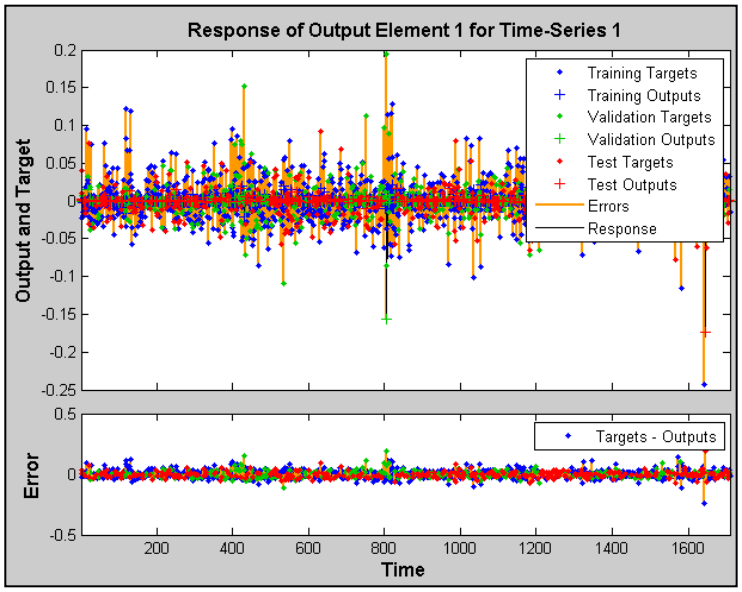


Figure 18: Plot of inputs, targets and errors against time

Lastly, the training set yielded a RMSE of 0.029467, with the validation set and testing set producing RMSE values of 0.030129 and 0.02679 respectively. These RMSE values appear to be in the range of forecastability of the other tools presented in this study. The testing set RMSE of 0.02679 is the lowest compared to all the forecasting tools presented and suggests that although the SAVI does not seem to be entirely forecastable, the best forecasting tool for the time series in question is a TDNN. Ladokhin (2009) reported RMSE values of 0.0919 and 0.0820 for the two neural networks used in that study.

Overall, the results suggest that the historical average model and the TDNN have better forecasting ability than all the other methods used in this study, with the historical average model only slightly outperforming the TDNN. The HAM model showing superior forecasting abilty goes against intuition, because of its simplicity in the way that volatility is forecasted. A longer sample period which would probably contain more shocks in the SAVI will probably distort the superior forecasting ability of the HAM, due to its simple approach, and inability to account for trends, non-linearity, or external factors affecting the SAVI.

It is interesting to note that the nonlinear models had lower RMSE values than the linear models (except for the HAM). This could be due to the fact that the SAVI is not random, and does display some non-linearity, as concluded from the BDS test, which was first published by Brock, Dechert, Scheinkman and Lebaron (1996) and is presented in Appendix D. The BDS test was conducted on the time series in question, but further tests were conducted on the linear models, all of which suggest that the models may be improved. This could be due to external factors affecting the volatility of the SAVI, other than historical values of the SAVI. This in turn allows room for artificial intelligence techniques to be explored more extensively, as these models allow for multiple inputs and factors which are assumed to affect the underlying variable to be forecast.

5. Conclusion

This chapter provides concluding remarks on the empirical analysis, with a discussion on avenues for further research.

5.1 Summary of Results

Table 16 below provides a summary of the results, based solely on a comparison of the RMSE values of the various forecasting tools explored. All of the RMSE values are fairly similar, with the EWMA model and the RMSE using the validation set of data for the TDNN producing the worst results. The best models can be concluded to be the TDNN, based on the RMSE of the testing set, and the HAM, although this method is considered a naïve approach and the result should be interpreted with caution. The mere fact that this simple analysis yielded positive results for the TDNN in terms of the best forecasting model of those compared, suggests that improvements to the model should produce even better results. Although a good forecasting model is said to have a RMSE close to zero, under the data constraints, the results could be improved greatly.

These results seem to suggest that the SAVI cannot be predicted, and that any guess at next period's volatility is just as good as a guess. The results should however be interpreted with caution as the sample used was fairly short due the SAVI only being introduced in 2007. Furthermore, no evidence exists of the SAVI being forecasted backwards, in order to obtain more data points. Furthermore, due to the returns of the SAVI being calculated by using the end of day price levels, a disparity may arise between the actual return of the SAVI at a particular point in time. In addition, based on the way that the SAVI is calculated, it may be possible that some of the options are infrequently traded, resulting in the closing prices of the SAVI being inaccurate measures of the actual closing prices at set points in time. The results may also improve if a larger sample is utilized, together with additional inputs used as potential forecasters.

Moreover, even though the analysis of the neural network suggests that the model could not be improved greatly; there still may be *caveats* to the construction and implementation of the

network. As the research surrounding this topic grows, further evidence may arise of specific rules and algorithms to use when making use of a neural network

Table 16: Summary of results

Model:	**RMSE**		
Historical Average (HAM)	0.025313		
Exponential Smoothing (ES)	0.029063		
Exponentially Weighted Moving Average (EWMA)	0.029529		
ARMA	0.028876		
ARCH	0.029018		
GARCH	0.028973		
	Training Set	**Validation Set**	**Testing Set**
Time-Delay Neural Network	0.029467	0.030129	0.02679

5.2 Recommendations for Further Research

As discussed previously, the fact that no consensus has been reached surrounding the best forecasting tool to use when forecasting volatility and stock prices, many avenues for further research in this field remain open. The growing popularity of artificial intelligence in finance adds to the many avenues of financial time series forecasting that can be expanded upon, as well as those that have yet to be explored.

Empirical analysis can be done by utilising different neural networks, with various architectures and neurodynamics. In addition, due to neural networks having the ability to solve complex problems, even when data is convoluted or contains missing values allows one to use additional factors besides historical prices or returns as inputs to the network. In particular, the daily open, close, high, and low values of the South African All Share Index can be used, together with other economic, technical, and fundamental factors as in Lawrence (1997). From a theoretical perspective, one can examine investor sentiment as being a driver of the SAVI, with proxies for this sentiment being used as an additional input in forecasting the SAVI.

Furthermore, combinations of forecasts can be used in order to obtain better forecasts, as proposed by Bates and Granger (1969) and Granger and Ramanathan (1984), who claim that combinations of forecasts from various models using different information may outperform the forecasts presented by any one particular model.

The direct link between market efficiency and the forecastability of the market can be explored further, whereby the ability to forecast volatility, with an investor being able to trade on such information, may lead one to hypothesise that the South African market in not entirely efficient. Furthermore, the concept of markets exhibiting cyclical efficiency, as proposed by Lo (2004) can also be examined, by testing for structural breaks in the data, as well as attempting to forecast volatility over longer or segmented periods of time.

Reference List

Abhyankar, A., Copeland, L., & Wong, W. (1997). Uncovering Non-linear Structure in Real-time Stock Market Indexes: The S&P 500, the DAX, the Nikkei 225 and the FTSE-100. *Journal of Business and Economic Statistics, 15*(1), 1-14.

Akigray, V. (1989). Conditional Heteroskedasticity in Time Series of Stock Returns: Evidence and Forecasts. *Journal of Business, 62*(1), 55-80.

Alexander, L., & Leigh, C. (1997). Covariance Matrices Used in Value at Risk Models. *Journal of Derivatives, 4*(1), 50-62.

Amiri, S., Von Rosen, D., & Zwanzig, S. (2009). The SVM Approach for Box-Jenkins Models. *REVSTAT Statisical Journal, 7*(1), 23-36.

Anderson, A., Carbone, R., Fildes, R., Hibon, M., Lewandowski, R., Makridakis, S., . . . Winkler, R. (1982). The Accuracy of Extrapolation (Time Series) Methods: Results of a Forecasting Competition. *Journal of Forecasting, 1*(1), 111-153.

Antonakakis, N., & Scharler, J. (2012). Volatility, Information and Stock Market Crashes. *Journal of Advanced Studies in Finance, 5*(1), 49-67.

Anwar, S., & Mikami, Y. (2011). Comparing Accuracy Performance of ANN, MLR and GARCH Models in Predicting Time Deposit Return of Islamic Bank. *International Journal of Trade Economics and Finance, 2*(1), 44-51.

Araujo, M., & Mare, E. (2006). Examining the Volatility Skew in the South African Equity Market Using Risk-neutral Historical Distributions. *Investments Analyst Journal, 64*(1), 15-20.

Archibald, B., & Koehler, A. (2003). Normaliztion of Seasonal Factors in Winters Methods. *International Journal of Forecasting, 19*(1), 143-148.

Atiya, F., El-Sherif, S., El-Shoura, S., & Shaheen, S. (1999). A Comparison Between Neural-Network Forecasting Techniques - Case Study: River Flow Forecasting. *IEEE Transactions on Neural Networks, 10*(2), 402-409.

Bardina, J., & Rajkumar, T. (2003). Training Data Requirement for a Neural Network to Predict Aerodynamic Coefficients. *AeroSense 2003: SOIE's 17th Annual International Symposium on Aerospace/Defense Simulation and Controls. 2.* Orlando: NASA Technical Reports Server.

Bates, J., & Granger, W. (1969). The Combination of Forecasts. *Operations Research Quarterly, 20*(1), 541-468.

Beckers, S. (1981). Standard Deviations Implied in Option Prices as Predictors of Future Stock Price Variability. *Journal of Banking Finance, 5*(1), 363-381.

Berk, J. (1995). A Critique of Size-Related Anomolies. *Review of Financial Studies, 8*(1), 275-286.

Bollerslev. (1986). Generalized Autoregressive Conditional Heteroskedasticity. *Journal of Econometrics, 31*(1), 307-327.

Bollerslev, T., Chou, R., & Kroner, K. (1992). ARCH Modelling in Finance. *Journal of Econometrics, 52*(1), 307-327.

Boudoukh, J., Richardson, M., & Whitelaw, R. (1997). Investigation of a Class of Volatility Estimators. *Journal of Derivatives, 4*(1), 63-71.

Bowerman, B., & O'Connell, R. (1979). *Time Series and Forecasting: An Applied Approach.* New York: Duxbury Press.

Box, G., & Jenkins, G. (1970). *Time Series Analysis: Forecasting and Control* (Vol. 1). San Francisco: Holden Day.

Boyd, M., & Kaastra, I. (1996). Designing a Neural Network for Forecasting Financial and Economic Time Series. *Neurocomputing, 10*, 215-236.

Brock, W., Dechert, J., Scheinkman, & LeBaron, B. (1996). A Test for Independence Based on the Correlation Dimension. *Econometric Review, 15*(1), 197-235.

Brock, W., Lakonishok, J., & LeBaron, B. (1992). Simple Technical Trading Rules and the Stochastic Properties of Stock Returns. *Journal of Finance, 41*(1), 1731-1764.

Brooks, C. (2008). *Introductory Economics for Finance* (Vol. 2). Cambridge: Cambridge University Press.

Brownlees, C., Engle, R., & Kelly, B. (2012). A Practical Guide to Volatility Forecasting Through Calm and Storm. *The Journal of Risk, 14*(2), 3-22.

Campbell, H. (1995). Predictable Risk and Returns in Emerging Markets. *Review of Economic Studies, 8*(1), 773-816.

Cancelliere, R., & Gemello, R. (1996). Efficient Training of Time Delay Neural Networks for Sequential Patterns. *Neurocomputing, 10*(1), 33-42.

Chen, A., Daouk, H., & Leung, M. (2000). Forecasting Stock Indices: A Comparisson of Classification and Level Estimation Models. *International Journal of Forecasting, 16*(1), 173-190.

Chiu, C., Lee, T., & Lu, C. (2009). Financial Time Series Forecasting Using Independent Component Analysis and Support Vector Regression. *Decision Support Systems, 47*(1), 115-125.

Clements, M., Franses, P., & Swanson, N. (2004). Forecasting Economic Time-series with Non-linear Models. *International Journal of Forecasting, 20*(1), 169-183.

Conrad, J., & KAul, G. (1988). Time Variation in Expected Returns. *Journal of Business, 61*(1), 409-425.

Crone, S., & Kourentzes, N. (2008). Forecasting High Freqeuency Time Series with Neural Networks - An Analysis of Modeling Challenges from Increasing Data Frequency. *International Conference on Data Mining*, (pp. 37-42). Las Vegas.

Day, T., & Lewis, C. (1992). Stock Market Volatility and the Information Content of Stock Index Options. *Journal of Economics, 52*(1), 267-287.

De Gooijer, J., & Hyndman, R. (2006). 25 Years of Time Series Forecasting. *International Journal of Forecasting , 22*(1), 443-473.

Deboeck, G. (1994). *Trading on the Edge: Neural, Genetic and Fuzy Systems for Chaotic Financial Markets.* New York: Wiley & Sons.

DeBondt, W., & Thaler, R. (1985). Does the Stock Market Overreact? *Journal of Finance, 40*(1), 793-805.

Dickey, D., & Fuller, W. (1979). Distribution of the Estimators for Autoregressive Time Series with a Unit Root. *Journal of the American Statistical Association, 74*(1), 427-431.

Dixit, G., Roy, D., & Uppal, N. (2013). Predicting India Volatility Index: An Application of Artificail Neural Network. *International Journal of Computer Applications, 70*(4), 22-30.

Dockner, E., Dorffner, G., & Schittenkopf, C. (2000). Forecasting Time-dependent Conditional Densities: A Semi-nonparametric neural network approach. *Journal of Forecasting, 19*(4), 355-374.

Donaldson, R., & Kamstra, M. (1997). An Artificia Neural Network-GARCH Model for International Stock Return Volatility. *Journal of Empirical Evidence, 4*(1), 17-46.

Donders, M., & Vorst, T. (1996). The Impact of Firm Specific News on Implied Volatilities. *Journal of Banking and Finance, 20*(1), 1447-1461.

Engle, R. (2004). Risk and Volatility: Econometric Models and Financial Practice. *The American Economic Review, 1*(1), 405-420.

Fair, R. (2000). *Events That Shook the Market.* Unpublished Thesis.

Fama, E. (1965). The Behaviour of Stock Market Prices. *Journal of Business, 38*(1), 34-105.

Fama, E. (1970). Efficient Capital Markets: A Review of Theory and Empirical Work. *Journal of Finance, 25*(1), 383-417.

Fama, E. (1991). Efficient Capital Markets II. *Journal of Finance, 46*(1), 1575-1617.

Ferson, W. (2007). Market Efficiency and Forecasting. In S. Satchell, *Forecasting Expected Returns in the Financial Markets* (pp. 1-15). Academic Press.

Figlewski, S. (1994). *Forecasting Volatility Using Historical Data.* Unpublished Thesis, New York University, Stern School of Business, New York.

Fleming, J. (1998). The Quality of Market Volaility Forecasts Implied by S&P 100 Index Option Prices. *Journal of Empirical Finance, 5*(1), 317-345.

French, K., Schwert, G., & Stampaugh, R. (1987). Expected Stock Returns and Volatility. *Journal of Financial Economics, 19*(1), 3-29.

Gahan, P., Mantri, J., & Nayak, B. (2010). Artificial Neural Networks: An Application to Stock Market Volatility. *International Journal of Engineering Science and Technology, 2*(5), 1451-1460.

Gardner, E. (1985). Exponential Smoothing: The State of the Art. *Journal of Forecasting, 4*(1), 1-28.

Geweke, J., & Meese, R. (1984). A Comparison of Autoregressive Univariate Forecasting Procedures for Macroeconomic Time Series. *Journal of Business and Economic Statistics, 2*(1), 191-200.

Gilchrist, W. (1976). *Statistical Forecasting.* John Wiley & Sons Ltd.

Granger, C. (1993). Strategies for Modelling Nonlinear Time Series Relationships. *Economic Record, 60*(1), 233-238.

Granger, C., & Poon, S. (2003). Forecasting Volatility in Financial Markets: A Review. *Journal of Economic Literature, 41*(1), 478-539.

Granger, C., & Poon, S. (2005). Practical Issues in Forecasting Volatility. *Financial Analysts Journal, 61*(1), 45-56.

Granger, C., & Timmermann, A. (2004). Efficient Market Hypothesis and Forecasting. *International Journal of Forecasting, 20*(1), 15-27.

Granger, W., & Ramanathan, R. (1984). Impreoved Methods of Combining Forecasting. *Journal of Forecasting, 3*(1), 197-204.

Guerrero, V., & Rosas, A. (1994). Restricted Forecasts Using Exponential Smoothing Techniques. *Internatioanl Journal of Forecasting, 10*(1), 515-527.

Hamilton, J., & Susmel, R. (1994). Autoregressive Conditional Heteroskedasticity and Changes in Regime. *Journal of Econometrics, 64*(1), 307-333.

Hayami,Y;Kuwahara,A;Matsumoto,K;Sakamoto,H. (2005). *IEICE Transactions on Fundamentals of Electronics , Communications and Computer Sciences, 88*(7), 1971-1978.

Hellstrom, T., & Holmstrom, K. (1998). *Predicting the Stock Market.* Unpublished Thesis, Malardalen University, Department of Mathematics and Physics, Vasteras, Sweden.

Hu, M., Patuwo, E., & Zhang, G. (1998). Forecasting with Artificial Neural Networks: The State of the Art. *14*(1), 35-62.

Hyndman, R. (2001). It's Time to Move From What to Why. *International Journal of Forecasting, 17*(1), 567-570.

Jensen, M. (1978). Some Anomalous Evidence Regarding Market Efficiency. *Journal of Financial Economics, 6*(1), 95-101.

Joseph, A., Koetze, A., & Oosthuizen, R. (2009). *The New South African Volatility Index.* Johannesburg. Retrieved from Johannesburg Stock Exchange Web site.

Juang, B., & Rabiner, L. (1993). *Fundamentals of Speech Recognition.* New Jersey: PTR Prentice-Hall, Inc.

Kaiser, M. (1994). *Time-Delay Neural Networks For Control.* Unpublished Thesis, University of Karlsruhe, Department of Computer Science, Karlsruhe.

Kim, H., & Shin, K. (2007). A Hybrid Approach Based on Neural Networks and Genetic Algorithms for Detecting Temporal Patterns in Stock Markets. *Applied Soft Computing, 7*(1), 569-576.

Ladokhin, S. (2009). *Forecasting Volatility in the Stock Market.* Unpublished Thesis, VU University Amsterdam, Faculty of Science.

Lamoureux, C., & Lastrapes, W. (1993). Forecasting Stock Return Variance: Toward an Understanding of Stochastic Implied Volatilities. *The Review of Financial Studies, 6*(1), 293-326.

Lawrence, R. (1997). *Using Neural Networks to Forecast Stock Market Prices .* Unpublished Thesis, University of Manitoba, Department of Computer Science, Manitoba.

Lewis-Beck, M. (2005). Election of Forecasting: Principles and Practice. *The British Journal of Politics and International Relations, 7*(1), 145-164.

Liu, Y., & Yang, M. (2012). The Forecasting Power of the Volatility Index in Emerging Markets: Evidence from the Taiwan Stock Market. *International Journal of Economics and Finance, 4*(2), 217-231.

Lo, A. (2004). The Adaptive Markets Hypothesis: Market Efficiency from an Evolutionary Perspective. *Journal of Portfolio Management, 30*(1), 15-29.

Lo, A., & MacKinlay, A. (1988). Stock Market Prices do not Follow Random Walks: Evidence From a Simple Specification Test. *Review of Financial Studies, 1*(1), 41-66.

Malkiel, B. (1992). Efficient Market Hypothesis. In M. Milgate, & P. Newman, *New Palgrave Dictionary of Money and Finance.* London: Macmillan.

Malliaris, M., & Salchenberger, L. (1996). Using Neural Networks to Forecast the S&P 100 Implied Volatility. *Neurocomputing, 10*(1), 183-195.

Mandelbrot, B. (1963). The Variation of Certain Speculative Prices. *Journal of Business, 36*(1), 394-419.

Marcellino, M. (2007). *A Comparisson of Time Series Models for Forecasting GDP Growth and Inflation.* Unpublished Thesis, European University Institute, Department of Economics, Milan.

Marucci, J. (2005). *Forecasting Stock Market Volatility with Regime-Switching GARCH Models.* Unpublished Thesis, University of California, Department of Economics, California.

Masters, T. (1993). *Practical Neural Network Recipes in C++.* New York: Academic Press.

Merton, R. (1980). On Estimating the Expected Return on the Market: An Exploratory Investigation. *Journal of Financial Economics, 8*(1), 323-361.

Miller, D., & Williams, D. (1999). Level-Adjusted Exponential Smoothing Modelling for Planned Discontinuities. *International Journal of Forecasting, 15*(1), 273-289.

Muth, J. (1960). Optimal Properties of Exponentially Weighted Forecasts. *Journal of the American Statistical Association, 55*(1), 299-306.

Nagpaul, P. (2005). *Time Series Analysis in WinIDAMS.* Unpublished Thesis, New Delhi.

Nelson, C., & Plosser, C. (1982). Trends and Random Walks in Macroeconomic Time Series. *Journal of Monetory Economics, 10*(1), 139-162.

Nelson, D. (1992). Filtering and Forecasting with Misspecified ARCH Models I: Getting the Right Variance with the Wrong Model. *Journal of Econometrics, 52*(1), 61-90.

Pissarenko. (2002). *Neural Networks for Financial Time Series Prediction: Overview Over Recent Research.* Unpublished Thesis, Department of Computer Science.

Poon, S. (2008, August 11). Historical Volatility Models. Retrieved August 2, 2013, from https://phps.portals.mbs.ac.uk/Portals/49/docs/spoon/HisVol.pdf.pdf

Qi, Z., & Zhang, G. (2005). Neural Network Forecasting for Seasonal and Trend Time Series. *European Journal of Operational Research, 160*, 501-514.

Rapach, D., & Zhou, G. (2013). Forecasting Stock Returns. In G. Elliot, & A. Timmermann, *Handbook of Economic Forecasting* (Vol. 2, pp. 328-383).

Roberts, S. (1982). A General Class of Holt-Winters Type Forecasting Models. *Management Science, 28*(1), 808-820.

Roll, R. (1984). A Simple Implicit Measure of the Effective Bid-Ask Spread in an Efficeint Market. *Journal of Finacne, 39*(1), 1127-1140.

Rozeff, P. (1984). Dividend Yields are Equtiy Premiums. *Journal of Portfolio Management, 1*(1), 68-75.

Sabbatini, M., & Linton, O. (1998). A GARCH Model of the Implied Volatility of the Swiss Market Index from Option Prices. *International Journal of Forecasting, 14*(1), 199-213.

Saleemah, A. (2012). A Multilayer Perceptron for Predicting Shear Strength of Reinforced Concrete Beams. *Journal of Civil Engineering and Construction Technology, 3*(2), 64-79.

Samouilhan, N., & Shannon, G. (2008). Forecasting Volatility on the JSE. *Investment Analysts Journal, 67*(1), 18-28.

Satchell, S., & Timmermann, A. (1995). On the Optimality of Adaptive Expectations: Muth Revisited. *International Journal of Forecasting, 11*(3), 407-416.

Sharpe, W. (1964). Capital Asset Prices: A Theory of Market Equilibtrium Under Conditions of Risk. *Journal of Finance, 19*(1), 425-442.

Shumway, R., & Stoffer, D. (2011). *Time Series Analysis and Its Applications With R Examples.* New York: Springer Science+Business Media.

Stuart, H. (1986). The Exponentially Weighted Moving Average. *Journal of Quality Technology, 18*(4), 203-210.

Sullivan, R., Timmerman, A., & White, H. (1999). Data-Snooping, Technical Trading Rule Performance and the Bootstrap. *Journal of Finance, 54*(1), 1647-1691.

Taylor, S. (1987). Forecasting the Volatility of Currency Exchange Rates. *Internatioanl Journal of Forecasting, 3*(1), 159-170.

Thiel, H. (1966). *Applied Economic Forecasting.* Chicago: Rand McNally.

Tse, Y. (1991). Stock Reurn Volatility in the Tokyo Stock Market. *Japan and the World Economy, 3*(1), 285-298.

Tse, Y., & Tung, S. (1992). Forecasting Volatility in the Singapore Stock Market. *Asia Pacific Journal of Management, 9*(1), 1-13.

Versace, M., & Wong, C. (2009). *Modeling Financial Time Series with Artificial Neural Networks.* Unpublished thesis, Boston University, Department of Cognitive and Neural Systems Center for Adaptive Systems, Boston.

Werbos, P. (1974). *Beyond Regression: New Tools for Prediction and Analysis in the Behavioural Sciences.* Unpublished Thesis, Harvard University.

White, H. (2000). A Reality Check for Data Snooping. *Econometrica, 68*(5), 1097-1126.

Wilamowski, M., & Yu, H. (2011). Levenberg-Marquardt Training. *The Industrial Electronics Handbook, 5*, pp. 12-27.

Yule, G. (1927). On the Method of Investigating Periodicities in Disturbed Series with Special Reference to Wolfer's Sunspot Nmbers. *Philosophical Transactions of the Royal Society of London, 226*, 267-298.

Zhang, G. (2001). An Investigation of Neural Network for Linear Time-series Forecasting. *European Computers and Operational Research, 28*(12), 1183-1202.

Appendix A1

This Appendix describes the minutia surrounding the forecasting of the ARCH model. The main diagnostic tests are presented in Chapter 4, which are useful for all the forecasting models used, and is not restricted to the ARCH model.

Table A1 below displays the main output from the ARCH model. The first half of the table below shows the standard output for the data generation process which is described by Equation (19) and is common to many of the forecasting tools used in this study. The lower part of the table details the coefficients, standard errors, z-statistics, and p-values for the coefficients of the conditional variance equation.

The conditional mean specification yields the following data generating process which underlies the daily returns on the SAVI:

$$y_t = -0.001026 - 0.174672y_{t-1} + 0.140598y_{t-2} - 0.089515y_{t-3} + 0.570276y_{t-4} + 10.203708\varepsilon_{t-1} - 0.194392\varepsilon_{t-2} + 0.074564\varepsilon_{t-3} - 0.509726\varepsilon_{t-4} \quad \text{(A1)}$$

with the conditional variance specification defined as:

$$h_t = 0.000642 + 0.283889{\varepsilon^2}_{t-1} \quad \text{(A2)}$$

It is interesting to note that the computations converged after 210 iterations, with not all of the coefficients corresponding to conditional mean specification being significant. The Maximum Liklihood (ML) estimates of the coefficients are, $\hat{\alpha}_0 = 0.000642$ and $\hat{\alpha}_1 = 0.283889$. Both of these coefficients have the correct sign, and are both significant at the 95% confidence level. The fact that $\hat{\alpha}_1$ is not very close to 1 indicates that volatility shocks are not very persistent.

Table A1: Analysis of ARCH model

Variable	Coefficient	Std. Error	z-Statistic	Prob.
C	-0.001026	0.000697	-1.472017	0.1410
AR(1)	-0.174672	0.292183	-0.597816	0.5500
AR(2)	0.140598	0.324989	0.432625	0.6653
AR(3)	-0.089515	0.279022	-0.320817	0.7483
AR(4)	0.570276	0.247898	2.300450	0.0214
MA(1)	0.203708	0.299429	0.680322	0.4963
MA(2)	-0.194392	0.333556	-0.582787	0.5600
MA(3)	0.074564	0.282658	0.263795	0.7919
MA(4)	-0.509726	0.247430	-2.060080	0.0394
Variance Equation				
C	0.000642	2.30E-05	27.87978	0.0000
RESID(-1)^2	0.283889	0.029388	9.660115	0.0000

R-squared	0.004159	Mean dependent var	-0.000165
Adjusted R-squared	-0.000600	S.D. dependent var	0.029087
S.E. of regression	0.029095	Akaike info criterion	-4.272737
Sum squared resid	1.417118	Schwarz criterion	-4.237258
Log likelihood	3606.509	Hannan-Quinn criter.	-4.259596
Durbin-Watson stat	2.063491		

Inverted AR Roots	.84	-.01+.83i	-.01-.83i	-.99
Inverted MA Roots	.83	-.02+.79i	-.02-.79i	-.99

Appendix A2

This Appendix outlines the analysis of the diagnostics before the actual forecasting is done when using the GARCH (1,1) model. As stated previously, the main diagnostic tests are presented in Chapter 4, which are useful for all the forecasting models used.

The table below displays the main output from the GARCH (1, 1) model. The first half of the table below shows the standard output for the data generation process which is described by Equation (19) and is common to many of the forecasting tools used in this study. The lower part of the table details the coefficients, standard errors, z-statistics, and p-values for the coefficients of the conditional variance equation, as was the case in Appendix A.

The conditional mean specification however, yields the following data generating process which underlies the daily returns on the SAVI:

$$y_t = -0.001263 + 0.892850y_{t-1} + 0.550582y_{t-2} - 0.817514y_{t-3} + 0.935557y_{t-4} + 0.880512\varepsilon_{t-1} - 0.531501\varepsilon_{t-2} + 0.825751\varepsilon_{t-3} - 0.965095\varepsilon_{t-4} \quad \text{(A3)}$$

with the conditional variance specification defined as:

$$h_t = 0.000105 + 0.156567{\varepsilon^2}_{t-1} + 0.725849h_{t-1} \quad \text{(A4)}$$

In this case, the computations converged after 51 iterations, with all of the coefficients corresponding to the conditional mean specification being significant. The Maximum Liklihood (ML) estimates of the coefficients are, $\hat{\alpha}_0 = 0.000105$, $\hat{\alpha}_1 = 0.156567$ and $\hat{\beta}_1 = 0.725849$. Both of these coefficients have the correct sign, and are both significant at the 95% confidence level. The fact that $\hat{\alpha}_1 + \beta_1 = 0.882416$, which close to 1, indicates that volatility shocks are quite consistent. This conclusion is contrary to the one made when examining the ARCH model.

Table A2: Analysis of GARCH model

Variable	Coefficient	Std. Error	z-Statistic	Prob.
C	-0.001263	0.000627	-2.016450	0.0438
AR(1)	0.892850	0.020912	42.69572	0.0000
AR(2)	-0.550582	0.033262	-16.55271	0.0000
AR(3)	0.817514	0.032038	25.51713	0.0000
AR(4)	-0.935557	0.020500	-45.63613	0.0000
MA(1)	-0.880512	0.017395	-50.61798	0.0000
MA(2)	0.531501	0.026122	20.34692	0.0000
MA(3)	-0.825751	0.024623	-33.53627	0.0000
MA(4)	0.965095	0.016329	59.10147	0.0000
Variance Equation				
C	0.000105	1.46E-05	7.142909	0.0000
RESID(-1)^2	0.156567	0.017263	9.069406	0.0000
GARCH(-1)	0.725849	0.028832	25.17534	0.0000
R-squared	0.007218	Mean dependent var		-0.000165
Adjusted R-squared	0.002473	S.D. dependent var		0.029087
S.E. of regression	0.029051	Akaike info criterion		-4.337466
Sum squared resid	1.412765	Schwarz criterion		-4.298762
Log likelihood	3661.978	Hannan-Quinn criter.		-4.323131
Durbin-Watson stat	2.028262			
Inverted AR Roots	.85-.51i	.85+.51i	-.41+.89i	-.41-.89i
Inverted MA Roots	.86-.51i	.86+.51i	-.42+.89i	-.42-.89i

Appendix B

The interested reader is referred to the following code which is used throughout the TDNN implementation:

```
function [newData1] = importfile(fileToRead1)
%IMPORTFILE(FILETOREAD1)
%  Imports data from the specified file
%  FILETOREAD1:  file to read

% Import the file
sheetName='Sheet1';
[numbers, strings] = xlsread(fileToRead1, sheetName);
if ~isempty(numbers)
    newData1.data =  numbers;
end
if ~isempty(strings)
    newData1.textdata =  strings;
end


% Solve an Autoregression Time-Series Problem with a NAR Neural Network
% Script generated by NTSTOOL
%
%
% This script assumes this variable is defined:
%
%   data - feedback time series.

targetSeries = tonndata(data,false,false);

% Create a Nonlinear Autoregressive Network
feedbackDelays = 1:2;
hiddenLayerSize = 10;
net = narnet(feedbackDelays,hiddenLayerSize);

% Choose Feedback Pre/Post-Processing Functions
% Settings for feedback input are automatically applied to feedback output
% For a list of all processing functions type: help nnprocess
net.inputs{1}.processFcns = {'removeconstantrows','mapminmax'};

% Prepare the Data for Training and Simulation
% The function PREPARETS prepares timeseries data for a particular network,
% shifting time by the minimum amount to fill input states and layer states.
% Using PREPARETS allows you to keep your original time series data
unchanged, while
% easily customizing it for networks with differing numbers of delays, with
% open loop or closed loop feedback modes.
[inputs,inputStates,layerStates,targets] = preparets(net,{},{},targetSeries);
```

```
% Setup Division of Data for Training, Validation, Testing
% For a list of all data division functions type: help nndivide
net.divideFcn = 'dividerand';  % Divide data randomly
net.divideMode = 'time';  % Divide up every value
net.divideParam.trainRatio = 60/100;
net.divideParam.valRatio = 20/100;
net.divideParam.testRatio = 20/100;

% Choose a Training Function
% For a list of all training functions type: help nntrain
net.trainFcn = 'trainlm';  % Levenberg-Marquardt

% Choose a Performance Function
% For a list of all performance functions type: help nnperformance
net.performFcn = 'mse';  % Mean squared error

% Choose Plot Functions
% For a list of all plot functions type: help nnplot
net.plotFcns = {'plotperform','plottrainstate','plotresponse', ...
  'ploterrcorr', 'plotinerrcorr'};


% Train the Network
[net,tr] = train(net,inputs,targets,inputStates,layerStates);

% Test the Network
outputs = net(inputs,inputStates,layerStates);
errors = gsubtract(targets,outputs);
performance = perform(net,targets,outputs)

% Recalculate Training, Validation and Test Performance
trainTargets = gmultiply(targets,tr.trainMask);
valTargets = gmultiply(targets,tr.valMask);
testTargets = gmultiply(targets,tr.testMask);
trainPerformance = perform(net,trainTargets,outputs)
valPerformance = perform(net,valTargets,outputs)
testPerformance = perform(net,testTargets,outputs)

% View the Network
view(net)

% Plots
% Uncomment these lines to enable various plots.
%figure, plotperform(tr)
%figure, plottrainstate(tr)
%figure, plotresponse(targets,outputs)
%figure, ploterrcorr(errors)
%figure, plotinerrcorr(inputs,errors)

% Closed Loop Network
% Use this network to do multi-step prediction.
% The function CLOSELOOP replaces the feedback input with a direct
% connection from the outout layer.
netc = closeloop(net);
[xc,xic,aic,tc] = preparets(netc,{},{},targetSeries);
```

```
yc = netc(xc,xic,aic);
perfc = perform(net,tc,yc)

% Early Prediction Network
% For some applications it helps to get the prediction a timestep early.
% The original network returns predicted y(t+1) at the same time it is given
y(t+1).
% For some applications such as decision making, it would help to have
predicted
% y(t+1) once y(t) is available, but before the actual y(t+1) occurs.
% The network can be made to return its output a timestep early by removing
one delay
% so that its minimal tap delay is now 0 instead of 1.  The new network
returns the
% same outputs as the original network, but outputs are shifted left one
timestep.
nets = removedelay(net);
[xs,xis,ais,ts] = preparets(nets,{},{},targetSeries);
ys = nets(xs,xis,ais);
closedLoopPerformance = perform(net,tc,yc)
```

Appendix C

This Appendix highlights a comparison of the Levenberg-Marquardt (LM) backpropagation algorithm with two other training algorithms. It is useful to understand the process behind the algorithm, as this affects the results of the neural network. The reasoning behind this choice of backpropagation algorithm is discussed earlier, with this appendix cementing prior statements for using the LM backpropagation algorithm.

Wilamowski and Yu (2011) illustrate the advantage of the LM algorithm in the following manner:

Three neurons in a multilayer perceptron network are utilised for training purposes, with the required training error being 0.01. A comparison of the convergent rate for each algorithm is done by tested one hundred trials with randomly generated weights; this can be seen in Figure C1 below.

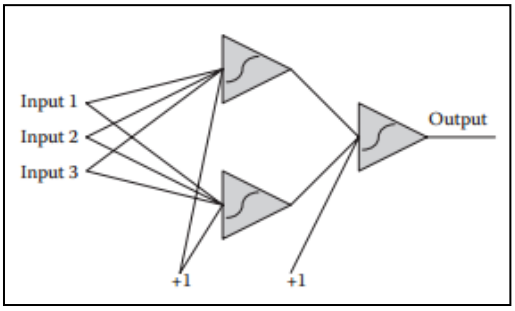


Figure C1: Three neurons in multilayer perceptron network (Source: Wilamowski and Yu (2011)

The training results are displayed in Table C1, which also provides the comparison. It can be seen that for the EBP algorithm, the larger the training constant (α) is, the faster and less stable the training process will be. Furthermore, it can be seen that the LM algorithm is a lot faster than the EBP algorithm, and more stable than the Gauss-Newton algorithm. Lastly, the Gauss-Newton method cannot converge at all for more multifaceted problems, and the EBP algorithm is said to

be inefficient when finding the solution, while the LM algorithm may result in more successful solutions.

Table C1: Comparison between three algorithms (Source: Wilamowski and Yu (2011)

Algorithms	Convergence Rate (%)	Average Iteration	Average Time (ms)
EBP algorithm ($\alpha = 1$)	100	1646.52	320.6
EBP algorithm ($\alpha = 100$)	79	171.48	36.5
Gauss–Newton algorithm	3	4.33	1.2
Levenberg–Marquardt algorithm	100	6.18	1.6

The LM algorithm is not without its flaws however. Wilamowski and Yu (2011) state that one problem with using this algorithm is that the Hessian matrix conversion has to be calculated for weight updating, with each iteration having many updates. This could remove the speed gained by using this algorithm, but is generally associated with more complex problems, such as image recognition problems.

Appendix D

This appendix displays the results of the BDS test for non-linearity. Brock, Dechert, Scheinkman and LeBaron (1996) designed the BDS test to test for non-random chaotic dynamics. The BDS test has the ability to detect remaining dependence as well as the presence of omitted non-linear structures.

In order to conduct the BDS test, a distance, ε, must first be chosen, where ε is a distance, whereby if the observations of the time-series and *iid*, then the probability of the distance between the two points being less than or equal to ε is constant.

This study makes use of a *fraction of pairs* to calculate ε, which ensures that a certain fraction of the number of pairs of points in the sample lie within ε of each other. Furthermore, the default value of 0.7 is used in calculating ε. The test statistic is calculated based the same ε, due to calculation efficiency. Due to the underlying series not being normal, and due to the possibility of the time series having an unusual distribution, bootstrapped p-values are also calculated.

Table D1 below displays the results of the BDS test. The Z-Statistics are all statistically insignificant, leading to a rejection of the null hypothesis that the time-series is *iid*. The value of ε is 0.037026, which is also used to calculate the BDS statistic, all of which are statistically insignificant at the 99% confidence level. These results indicate that there could be hidden nonlinearity in the time series, which would justify using nonlinear methods of forecasting.

Table D1: BDS test for non-linearity

Dimension	BDS Statistic	Std. Error	z-Statistic	Normal Prob.	Bootstrap Pr...
2	0.011969	0.002170	5.516047	0.0000	0.0000
3	0.025936	0.003443	7.532134	0.0000	0.0000
4	0.031979	0.004095	7.810067	0.0000	0.0000
5	0.033688	0.004262	7.904384	0.0000	0.0000
6	0.034506	0.004105	8.406727	0.0000	0.0000
Raw epsilon		0.037026			
Pairs within epsilon		2001185.	V-Statistic	0.703165	
Triples within epsilon		2.59E+09	V-Statistic	0.538989	