Forecasting changes in the South African volatility index: A comparison of methods

Ushir HARRILALL[1], Yudhvir SEETHARAM[2]

[1]*University of the Witwatersrand, South Africa*

[2]*University of the Witwatersrand, South Africa, yudhvir.seetharam@wits.ac.za*

Abstract. Increased financial regulation with tougher capital standards and additional capital buffers has made understanding volatility in financial markets more imperative. This study investigates 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 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. 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. 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.

Keywords: forecasting, volatility index, neural networks, time series, emerging markets

1 Introduction

Volatility prediction has become a crucial task in the appraisal of assets and risk management. 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.

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 to explore interrelationships in financial markets; in particular, Artificial Neural Networks. 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. This study analyses and examines the use of neural networks as well as traditional and non-linear methods as a forecasting tool. Specifically, a comparison is done between traditional forecasting methods and a Time-Delay Neural Network (TDNN) with regard to the various models' ability to forecast volatility.

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. The SAVI was chosen to explore the impact of forecasting methods on an emerging market volatility index. We examine the null hypothesis that the SAVI cannot be forecasted using neural networks against its alternative; along with examining whether a neural network is not a better modelling technique than traditional forecasting techniques. These traditional techniques are selected from the literature and each subsequent method can be seen as an "improvement" over its predecessor. For example, we consider basic moving average methods, improved by conditional volatility models, improved by non-linear models. Further, the choice of particular neural networks to use was governed by those that are more apt to time series data. We compare our models using a traditional RMSE criterion (any alternate criterion may well be used) and make the assumption that the changes to the SAVI are not random. This assumption allows us to create a foundation upon which we can model the SAVI.

The article proceeds with a review of forecasting methods and a background of the SAVI, followed by a description of the models and data used, ending with an analysis of the results and concluding remarks.

2 Literature review

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.

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.

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. 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. The author's empirical analysis suggests superior forecastability.

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). 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. This study adopts a similar methodology of comparison and is discussed later.

The complex nature of markets, as well as the discovery of markets exhibiting non-linear movements (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.

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.

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.

The South African Volatility Index – An Introduction

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 1 below. As can be seen from Figure 1 below, the SAVI shows visual signs of clustering and potential signs of mean reversion. Given the objective of this paper - to determine the best forecasting method for the SAVI - these two statistical considerations provide a possible hindrance to the objective. Stein (1989) find that option traders in the US overreact to implied volatility measures when new information arrives for short term option contracts. This overreaction would result in some form of clustering. The author shows that while these overreactions are statistically significant, they have a small economic impact. We rely on this finding and therefore make this assumption in our paper. Naturally, an avenue for future research would be to test this assumption on the SAVI.

Further, the author postulates that while mean reversion is typically present in volatility, this statistical anomaly is of concern when considering the implied volatility of long-dated options. In this scenario, the economic significance of overreaction would be much more apparent. In the context of the SAVI, the option maturities used are typically short-term in nature. Therefore, the anomaly of mean reversion (and its economic significance, if mispriced), is of minor importance.

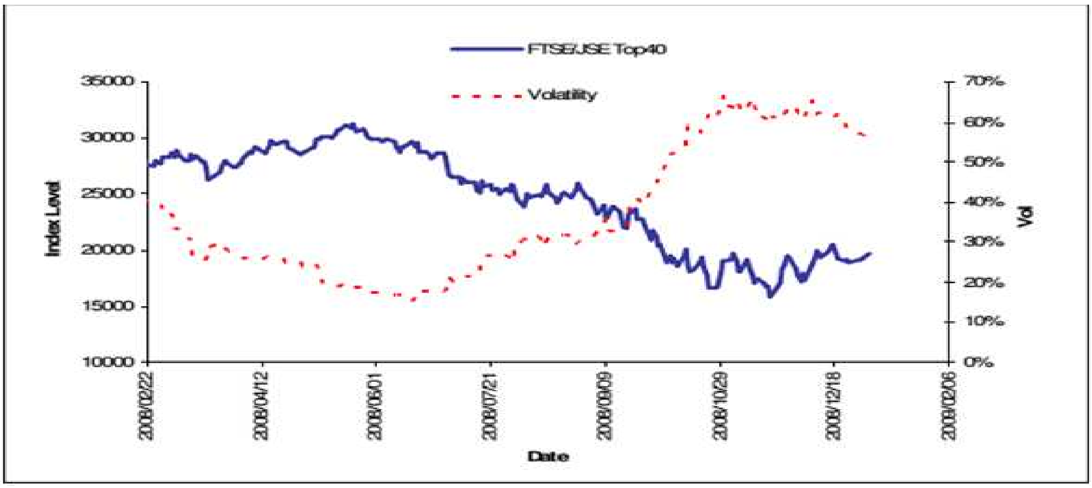

Figure 1 FTSE/JSE Top40 index level and its volatility

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

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

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.

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.

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 options are illiquid, with prices not always being transparent (Araujo & Mare, 2006).

3 Research methodology

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 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. Aboura and Wagner (2014) show that extreme asymmetric volatility can have a significant role in explaining periods of market downturns. Our sample period was chosen both out of necessity (due to the SAVI calculation being restated); and to also correspond to a full business cycle. It is quite possible that the results could be mired with the presence of volatility feedback (as per Aboura and Wagner, 2014), in that large market downturns are a consequence of volatility feedback and rational asset pricing behaviour. We require a larger sample period to fully investigate this phenomenon; and that relies on re-engineering the SAVI calculation before it was restated in 2010. 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. 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.

Historical Volatility Models

a) 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:

$$\hat{\sigma}_{t+1} == \frac{1}{t}\sum_{i=1}^{t}\sigma_i \tag{2}$$

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

b) Simple Exponential Smoothing (SES)

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

$$\sigma_t = (1-\alpha)\sigma_{t-1} + \alpha\hat{\sigma}_{t-1} + \zeta_t \tag{3}$$

$$\hat{\sigma}_{t+1} = (1-\alpha)\sigma_t + \alpha\hat{\sigma}_t \tag{4}$$

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 previously, 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.

c) Exponentially Weighted Moving Average (EWMA)

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

$$\sigma^2_{t+1|t} = \sum_{j=1}^{n} w_j\left(R_{t-j+1} - \overline{R}_t\right)^2 \text{ where } w_j = (1-\lambda)\lambda^{j-1} \tag{5}$$

The EWMA is seen as a more sophisticated method compared to the SMA, in that it assigns higher weightings to more recent observations (whereas the SMA assigns equal weightings to all observations, irrespective of how much time has passed). Shumway and Stoffer (2011) argue that the ARIMA (0,1,1) model leads to an EWMA model. Furthermore, the authors define a smoothing parameter, λ, 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 Holt-Winters 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 Holt-Winters three parameter model. The Holt-Winters 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 Holt-Winters 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.

Autoregressive and Heteroskedastic Models

a) 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:

$$\hat{\sigma}_{t+1} = \sum_{i=1}^{p} \lambda_i \sigma_{t+1-i} + \sum_{i=1}^{q} \gamma_i \zeta_{t+1-i} \quad (6)$$

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)

b) 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. The ARCH model is described by the following formulae as presented by Ladokhin (2009):

$$r_t = \mu + \varepsilon_t \quad (7)$$

$$\varepsilon_t = \sqrt{h_t z_t} \quad (8)$$

$$h_t = \overline{\omega} + \sum_{j=1}^{q} \alpha_j \varepsilon_{t-j}^2 \quad (9)$$

where r_t is the return at time t, μ is the mean return, ε_t are the residuals, $\overline{\omega}$ and α_j are parameters of the model, h_t is the conditional variance with z_t ~ *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, $\overline{\omega}$>0 and α_j≥0 (Ladhokhin, 2009).

c) 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:

$$h_t = \overline{\omega} + \sum_{i=1}^{p} \beta_i h_{t-i} + \sum_{j=1}^{q} \alpha_j \varepsilon_{t-j}^2 \quad (10)$$

where the parameter β_i>0.

Neural Network Model Construction

a) 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 (Pissarenko, 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

$$Oj = transfer \sum(x_{ji} w_{ji}) \quad (11)$$

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.

b) 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.

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 (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.

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.

c) 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 non-linear function. This algorithm is one of the most popular tools for non-linear 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).*

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). 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 in transforming the input data. 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:

$$transfer(x) = \tanh(x) = \frac{\sinh(x)}{\cosh(x)} = \frac{e^{x} - e^{-x}}{e^{x} + e^{-x}} \quad (12)$$

d) 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.

e) 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.

4 Empirical results

Examining the Univariate Properties

We examine the data for evidence of independence (via a correlogram and Q-statistics), normality (via a Jaque Bera test) and for stationarity (via the Augmented Dickey Fuller and Phillips-Peron test). Our results are not reported here in detail and are available upon request. We find that there is evidence of independence in our data as the autocorrelation coefficients drop to zero quickly, and our data 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 non-linear processes which allow successive index changes to be linked through the variance.

Second, the Jarque-Bera test statistic, which jointly examines the skewness and kurtosis to formulate a joint test statistic leads us to reject the null hypothesis of normality at a 99% confidence level, as the *p-value* is less than 0.01. 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.

Third, this study makes use of the Augmented Dickey-Fuller test with a constant and a linear trend. The results 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 suggest that one would reject the null hypothesis of the return series having a unit root at the 99% confidence level.

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:

$$y_t = \delta + \sum_{i=1}^{p} \emptyset_i y_{t-i} + \varepsilon_t + \sum_{j=1}^{q} \theta_j \varepsilon_{t-j} \quad (13)$$

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:

$$y_t = \delta + \emptyset_1 y_{t-1} + \emptyset_2 y_{t-2} + \emptyset_3 y_{t-3} + \emptyset_4 y_{t-4} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \theta_3 \varepsilon_{t-3} + \theta_4 \varepsilon_{t-4} \quad (14)$$

Using the minimum AIC criterion, it was found that the appropriate lag order was 4. Table 1 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 1 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

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.

When examining Table 2, no root lies outside the unit root circle, indicating that the ARMA model is invertible and stationary. Figure 2 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.

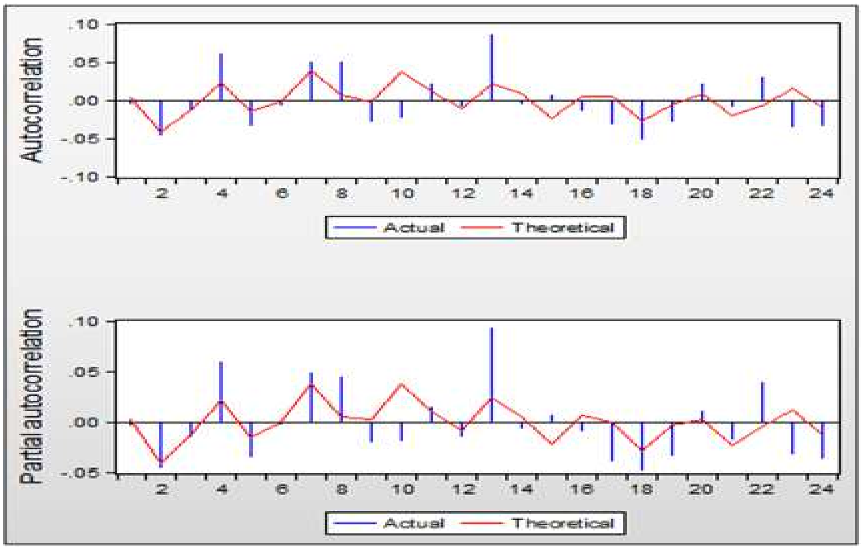

Figure 2 Graphical view of the actual and ARMA model correlogram

Table 2 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.	

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 3 displays the impulse response function of the ARMA model. 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 3.

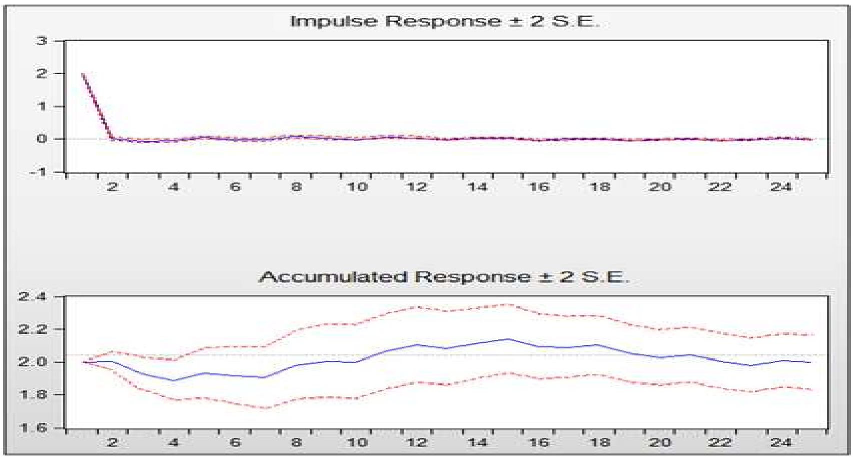

Figure 3. Impulse response function of ARMA model

Analysing the residuals of the model (not displayed here), 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.

Testing for ARCH Effects

Table 3 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 3 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 3 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

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.

Exponential Smoothing

The parameters of the ES model are obtained by minimising the sum of squared errors. 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 4 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. 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 4 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

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 α=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 5 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 5 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

ARMA Model

Table 6 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 Inequality Coefficient (Thiel's U) 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 6 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

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 7 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 7 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

GARCH

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

Table 8 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 8 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

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 (60% for training, 20% for validation and 20% for forecasting). 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)*. The network is generated and trained in open loop form as 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 4 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.*

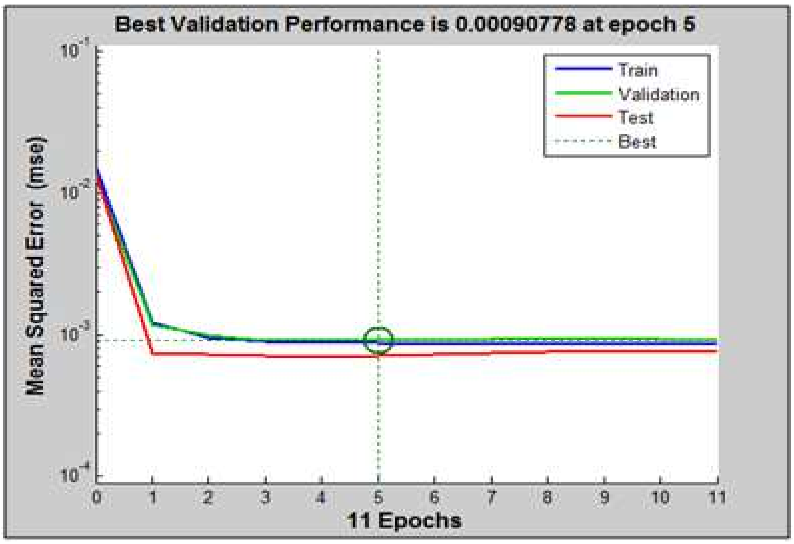

Figure 4 Best validation performance

One now turns to examine the error terms of the network. 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. While unreported, 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.

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 non-linear 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 (Brock, Dechert, Scheinkman and LeBaron, 1996). This test for non-linearity examines the distribution of data at each quartile for linear behaviour. The results of the test (not presented here), show that the SAVI does indeed exhibit non-linear behaviour at each quartile. 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

Table 9 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.

Given our findings that the SAVI cannot be forecasted, we now explore some potential implications of our result. From an economic or theoretical perspective, the absence of predictability of the SAVI can point towards a market where sentiment (as given by option volatility) cannot be reliably predicted. While it is tempting to link our results to comment on market efficiency, it is tenuous to compare a lack of prediction on market sentiment to whether equity prices reflect all available information. Further, the SAVI itself is not a tradable product and is only considered an indicator of market sentiment. However, if one were interested in trading volatility-based assets, the only security on offer to South African investors is a variance futures contract. Lastly, given our lack of predictability, a trading strategy cannot be relied upon to generate consistent, abnormal profits.

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 9 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

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. 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.

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.

6 References

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), pp. 1-14.

Aboura, S., & Wagner, N. (2012). Extreme asymmetric volatility, leverage, feedback and asset prices. *Working Paper, University of Paris Dauphine and Passau University*.

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

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), pp. 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), pp. 15-20.

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.

Bollerslev, T. (1986). Generalized Autoregressive Conditional Heteroskedasticity. *Journal of Econometrics, 31*(1), pp. 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), pp. 63-71.

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,* pp. 215-236.

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

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

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

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

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), pp. 1451-1460.

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

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

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

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

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

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

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

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

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

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), pp. 199-213.

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

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

Stein, J. (1989). Overreactions in the options market. *The Journal of Finance,44*(4), pp. 1011-1023.

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

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

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

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

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

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