Courant Research Centre
'Poverty, Equity and Growth in Developing and Transition Countries: Statistical Methods and Empirical Analysis'

Georg-August-Universität Göttingen
(founded in 1737)

Discussion Papers

No. 203

Comparing the market risk premia forecasts in JSE and NYSE equity markets

Leoni Eleni Oikonomikou

April 2016

Wilhelm-Weber-Str. 2 · 37073 Goettingen · Germany
Phone: +49-(0)551-3914066 · Fax: +49-(0)551-3914059

Email: crc-peg@uni-goettingen.de Web: http://www.uni-goettingen.de/crc-peg

Comparing the market risk premia forecasts in JSE and NYSE equity markets

by Leoni Eleni OIKONOMIKOU[1]

First version: March 2015

Revised version: November 2015

Corrected version: March 2016

[1] Research Associate and PhD Candidate at Georg-August University of Goettingen, Germany.
Contact address: Platz der Göttinger Sieben 5, 37073, Göttingen, Germany
e-mail: leoni-eleni.oikonomikou@wiwi.uni-goettingen.de , le.oikonom@gmail.com
I would like to thank the DFG foundation, RTG 1723 for the provided financial support along the preparation of this research paper

Comparing the market risk premia forecasts in JSE and NYSE equity markets

Abstract:

This paper examines the evidence regarding predictability in the market risk premium using artificial neural networks (ANNs), namely the Elman Network (EN) and the Higher Order Neural network (HONN), univariate ARMA and exponential smoothing techniques, such as Single Exponential Smoothing (SES) and Exponentially Weighted Moving Average (EWMA).

The contribution of this paper is the inclusion of the South African market risk premium to the forecasting exercise and its direct comparison with US forecasting results. The market risk premium is defined as the expected rate of return on the market portfolio in excess of the short-term interest rate for each market. All data are taken from January 2007 till December 2014 on a daily basis.

Elman networks provide superior results among the tested models in both insample and out-of sample periods as well as among the tested markets. In general, neural networks beat the naive benchmark model and achieve to perform better than the rest of their linear tested counterparts.

The forecasting models successfully capture patterns in the data that improve the forecasting accuracy of the tested models. Therefore, they can be applied to trading and investment purposes.

Keywords: forecasting performance, market risk premium, South African stock market, US stock market

JEL Classification: C45, C52, G15, G17

1. Introduction

In the recent days, the improvement of information technology has led investors to become more dependent on advanced computer and communication technicalities in order to benefit from a broader range of investment choices. Artificial neural networks are considered as a promising area of research within this financial environment. One of their most important properties is that they can theoretically approximate any non-linear continuous function on a compact domain to any designed degree of accuracy (Cybenko, 1989).

Some other challenges that portfolio managers, investors and large financial institutions confront, concern the accurate forecasting of future moves in stock markets. In their quest to forecast the equity markets, they make the assumption that future occurences are based in part on present, past events and data. Besides, financial time series are very "noisy" and their signals are quite difficult to be forecasted (Abu-Mostafa and Atiya, 1996).

In this paper, I aim to forecast the market risk premium for two financial markets. The market risk premium is defined as the expected return over a risk-free rate that investors request from the market in order to bear the increasing risk on investing in this specific market. It is considered as an important component while deciding on whether to invest in an emerging compared to a developed market, in a risky stock relative to a riskless bond and/or in stocks or bonds compared to other asset classes such as foreign exchange, commodities, investment funds and so on.

Of further interest is the calculation of the cost of capital for a corporation, alternatively referred to as Weighted Average Cost of Capital (WACC). Corporate finance professionals seek to value one's company (its required cost of equity, in particular). In essence, daily calculation updates of the market risk premium, i.e. the premium of the aggregate stock market, multiplied by beta (the sensitivity of the company towards risk), are necessary for corporate valuation purposes. For further information on how to calculate and estimate the market risk premium refer to Damodaran (2013). In order to get to know better about the debate on whether econometric models yield better forecasts of equity premium than the historical average refer to Spiegel (2008).

Economists also care about the level of the market risk premium in long-term horizons since it can be accounted as an indicator of macroeconomic stability of the country. The market risk premium can also be seen as an indicator of the performance of the stock market relative to the bond market. Furthermore, in the real sector, emerging market economies use various policy tools, such as capital controls or sterilized reserve accumulation, in order to smooth the fluctuations in risk premia on their economy. Their ability to issue safe assets raises emerging markets economy welfare but decreases, at the same time, that of the advanced economy.

During the past 15 years, improved fundamentals and the low yields in advanced economies have attracted a broad range of investors to increase their share of investment in the financial assets of emerging market economies. This has fostered and deepened the development of local financial markets and of new asset classes, like local-currency denominated sovereign debt and expanded the GDP growth rate. In the same vein, the relative role of cross-border bank lending has declined and, within portfolio flows, fixed-income flows have gained in significance compared with equity flows. Globalization of financial markets has also promoted the growing importance of globally operating mutual funds that do not only focus on emerging market economies but take into account a broader investment scale of countries (IMF, 2014).

It is widely known that the foreign presence in local emerging markets is quite significant in terms of size and the one that promotes the development of these markets and their economies alike. For instance, a well established stock market can channel the savings into investment and can also provide an alternative funding for local businesses, beyond the local banking system and the level of domestic inflation.

My methodological approach aims at comparing various forecasting methods namely univariate ARMA and EWMA methods with the so-called multivariate neural networks. This study is interesting because not only does it cover the univariate forecasting models, widely used in the financial and business industry, but also accounts for more variables inclusion in the forecasting exercise via the neural networks application. The application of neural networks is important because as mentioned later in the paper, they are self-adaptive models, driven by the data. Further, no a priori assumption about the statistical distribution of the data or any particular model specification is needed, during the model building process. In addition, they are non-linear models that enable better forecasting of complex data patterns, in comparison with well-known linear methods such as ARIMA.

The remainder of the paper is organized as follows. Section 2 summarizes relevant literature. Section 3 introduces the methods used in this study and Section 4 provides an extensive description of the data. Finally, Section 5 presents the empirical results and Section 6 concludes.

2. Literature Review

Several studies tried to forecast the equity/market risk premium in the US financial market. The capabilities of general equilibrium asset pricing models to forecast the market risk premium have been questioned in the past (Welch and Goyal, 2006). In their study, Welch and Goyal prove that a significant number of economic variables with in-sample predictive ability for the equity/market risk premium fail to achieve consistent out-of-sample forecasting gains relative to the historical average. In his study, Spiegel (2008) sets the following research questions: Can our empirical models accurately forecast the equity premium any better than the historical mean? Does forecasting via the widely known empirical models give us more accurate than what we would get by simply beating the historical mean?

Rapach *et al.* (2010) propose the combination of individual forecasts. They argue that combining forecasts incorporates information from various economic variables while it substantially diminishes forecast volatility, and this combination makes the forecasts more realistic. Neely *et al*. (2014) combine macroeconomic with technical indicators variables which provide additional information over the business cycle. More precisely, technical indicators better identify the decline in the equity risk premium near business-cycle peaks, while macroeconomic variables pick up the typical rise in the equity/market risk premium near cyclical troughs. They also demonstrate that this combination successfully captures any countercyclical fluctuations in the equity/market risk premium that may appear.

My research aims to provide an alternative view (beyond the scope of classic general equilibrium modelling) on forecasting the market risk premium in short time horizons. The built-in forecast model is a customized one and its inputs are selected based on the correlation with the target time series, namely the market risk premium.

In the past few decades, furthering with the evolution in computer technology, many studies have applied soft computing techniques. Artificial Neural Networks (ANNs) have become popular and have been applied in several financial prediction problems. They are presented as a nonlinear nonparametric alternative to traditional linear-based forecasting methods (G. Zhang *et al.*, 1998).

Qi (1999) investigates evidence in nonlinear predictability for the excess returns on S&P 500 by applying recursive neural network. His study differs from mine in two directions. First, he benchmarks his neural network model to a linear regression model, whereas I benchmark it to a simple naive model. Second, he tests his results based on trading strategies for switching portfolios, whereas my study does not include testing of the forecast model with trading strategies.

The vast majority of studies concerning S&P 500 and more broadly the US stock market incorporate trading strategies within the customized forecast model. Chenoweth and Obradovic (1996) suggest a stock market prediction system that is consisted of two components namely, two neural networks and a decision rule base. Their results on daily and monthly experiments indicate that the dual neural network system outperforms the single one by providing higher returns with fewer trades. Motiwalla and Wahab (2000) propose a switching rule accounting the one-step ahead prediction of returns and exploring the establishment of investment positions in stocks or T-Bills. Moreover, they explore the economic significance of their investment strategy by incorporating transaction costs into the simulation of trading strategies. Their results concerning the ANNs are promising. Tsaih *et al.* (1998) propose a hybrid artificial intelligence (AI) in order to implement trading strategies in the S&P 500 stock index futures market. This approach incorporates the rule-based systems (RBS) method and the neural networks (NNs) method in order to predict the direction of daily price changes of the index. Their empirical results confirm the superiority of this hybrid approach to a passive buy-and-hold investment strategy.

The literature on financial forecasting with machine learning methods is also vast. Mirmirani and Li (2004) apply the NeuroGenetic Optimizer Software to the NYMEX database of daily gold cash price which covers the following period of data: 12/31/1974-12/311998. Several methods have been applied in order to predict gold price movements such as the backpropagation NNs with genetic algorithms. Their empirical findings suggest that prices in the past significantly affect the future gold prices confirming the fact of short-term dependence in gold price movements. Olson and Mossman (2003) compare ordinary least squares (OLS) and linear regression (logit) methods with NN forecasts of one-year Canadian stock returns. Their empirical findings suggest that backpropagation NNs, outperform the best regression alternatives for both point estimation and in classifying firms. This superiority in performance of NNs is also interpreted to greater profitability with various trading rules. De Faria *et al.* (2009) compare the forecasting performance of NNs and the adaptive exponential smoothing method in the Brazilian stock market. They find that NNs perform better than the adaptive exponential smoothing method and conclude that the former can be used to develop investment strategies. Kara *et al.*(2011) apply ANNs and support vector machines (SVMs) in the task of forecasting the direction of change in the daily Istanbul Stock Exchange (ISE) National 100 Index, using ten technical indicators. Their results confirm the superior average performance of ANNs over the SVMs models. Dunis *et al.* (2014) successfully apply gene expression and integrated genetic programming algorithms in modeling and trading the ASE 20 Greek Index and their results outperform commonly existing methods. Roh (2007) suggests hybrid models with NN and time series models for forecasting the volatility of stock price index in both the deviation and direction.

Witkowska and Marcinkiewicz (2005) results favor the combination of technical analysis and ANNs in order to exploit the possibility of profiting from trading in Warsaw Stock Exchange Futures Market. Kiani and Kastens (2008) model relationships between futures contracts on exchange rates for British pound (BP), Canadian dollar (CD) and Japanese yen (JY) against US dollars, using linear models, feed forward artificial neural networks (ANN) and three versions of recurrent neural networks for predicting exchange rates against the US dollar. Walczak (1999) presents evidence that the Singapore stock market is affected by external signals and exploits trading advantages from these signals. He applies NNs on trading market indices and compares his results with Dow Jones market index. Results suggest that these signals successfully improve the forecasting accuracy on the Singapore DB50 index but have little or almost no effect on forecasts for Dow Jones Industrial Average Index.

Hæke and Helmenstein (1998) construct an index of Austrian Initial Public Offerings (IPOX) isomorphic to the Austrian Traded Index (ATX). They investigate the time trend properties and comovements between the two indices. This relationship is then used in order to build a NN and a linear error-correction forecasting model for the IPOX and benchmark a trading scheme on each forecast.

Kumar and Haynes (2003) apply an ANN in order to forecast credit rating in India. Their results demonstrate increasing speed and efficiency of the rating process when ANN model is applied. They also confirm the superior performance of ANNs over their benchmark models.

Bekiros and Georgoutsos (2008) apply Elman networks to develop a trading algorithm for the NASDAQ composite index. Their results demonstrate that the best performing model is the one that accounts for conditional volatility.

Thinyane and Millin (2011) apply genetic algorithms (GAs) and artificial neural networks (ANNs) in order to generate signals by technical trading tools which are optimized for maximum profit. The final result is an autonomous intelligent trading system which is proven to be profitable based on data of 10 currencies spanning over a five year period. Further, the profit margins are statistically significant and statistically significantly more profitable than other no-risk investment strategies.

3. Methodology

3.1 Benchmark naive model

The naive model is commonly applied within the forecasting literature in order to predict the future. It assumes that all future forecasts are equal to the last observed value of the time series (the market risk premium, in our case). Therefore, the model has the following form:

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

where:

- Y_t is the current rate of return at time t
- Y_{t+1} is the forecast rate of return at time $t+1$

3.2 Exponential Smoothing Methods

Exponential Smoothing techniques attach larger weights to more recent observations as compared to observations longer in the past.

The single exponential smoothing model takes the following form:

$$Y_{t+1} = ay_t + (1-a)Y_t \qquad 0 < a < 1 \qquad (2) \qquad \text{and} \quad Y_{t+1} = Y_t + a\varepsilon_t \qquad (3)$$

where

- y_t actual observation
- Y_t smoothed value
- ε_t is the forecast error for period t
- a smoothing parameter that takes the value 0.2 in this setting

Put it simply, the new forecast is the old one in addition to the adjustment for the error that occurred in the last forecast.

3.3 Exponentially Weighted Moving Average (EWMA)

The Exponentially Weighted Moving Average (EWMA) takes averages of the data in such a way that gives less weight to data further in the past. Depending on the choice of the weighting factor λ, this constant determines the depth of the memory of EWMA process. The EWMA statistic takes the following form:

$$EWMA_t = \lambda Y_t + (1-\lambda)EWMA_{t-1} \qquad \text{for} \qquad t = 1, 2, ..., n \tag{4}$$

where:

- $EWMA_0$ is the mean of historical data (target)
- Y_t is the observation at time t
- n is the number of observations to be monitored including $EWMA_0$
- $0 < \lambda \leq 1$ is a constant that determines the depth of memory of the EWMA.

The parameter λ determines the rate at which previous data enter into the calculation of the EWMA statistic. A value of $\lambda = 1$ implies that only the most recent measurement influences the EWMA. The value of λ is usually set between 0.2 and 0.3. In this setting $\lambda = 0.2$.

3.4 Autoregressive Moving Average Model (ARMA)

ARMA models include two components, the autoregressive (AR) and the moving average (MA) one. Therefore, for a time series Y_t, the level of its current observations depends on the level of its lagged observations as well as the shocks that happened at time t and the lagged shocks at times $t-1$.

The combined ARMA process is given by the following form:

$$Y_t = \varphi_0 + \varphi_1 Y_{t-1} + \varphi_2 Y_{t-2} + ... + \varphi_p Y_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - ... - \theta_q \varepsilon_{t-q} \quad (5)$$

where:

- Y_t is the dependent variable at time t
- $Y_{t-1}, Y_{t-2}, ..., Y_{t-p}$ are the lagged dependent variables
- $\varphi_0, \varphi_1, ..., \varphi_p$ are the regression coefficients
- ε_t is the error or shock term
- $\varepsilon_{t-1}, \varepsilon_{t-2}, ..., \varepsilon_{t-q}$ are the previous values of the error or shock term
- $\theta_1, \theta_2, ..., \theta_q$ are the error weights

Taking into consideration the insample correlogram as well as the AIC model selection criteria, an ARMA(4,4) model was chosen for South Africa and an ARMA (5,4) model for the US. All the coefficients are statistically significant (except from the constant in the South African case and the MA(1) in the US case).

The two ARMA models for South Africa and US, applied in this study are specified and the output results are cited in the Appendices A.1- A.2:

- $$Y_t = 0.014963 - 0.772943 Y_{t-1} + 0.203301 Y_{t-2} - 0.730878 Y_{t-3} - 0.969394 Y_{t-4} - 0.797358 \varepsilon_{t-1} + 0.166009 \varepsilon_{t-2} - 0.723628 \varepsilon_{t-3} - 0.940598 \varepsilon_{t-4} \quad (6)$$

- $$Y_t = 0.014178 + 0.036229 Y_{t-1} + 0.660882 Y_{t-2} - 0.445786 Y_{t-3} - 0.539741 Y_{t-4} - 0.16352 Y_{t-5} + 0.165128 \varepsilon_{t-1} + 0.726791 \varepsilon_{t-2} - 0.579933 \varepsilon_{t-3} - 0.530582 \varepsilon_{t-4} \quad (7)$$

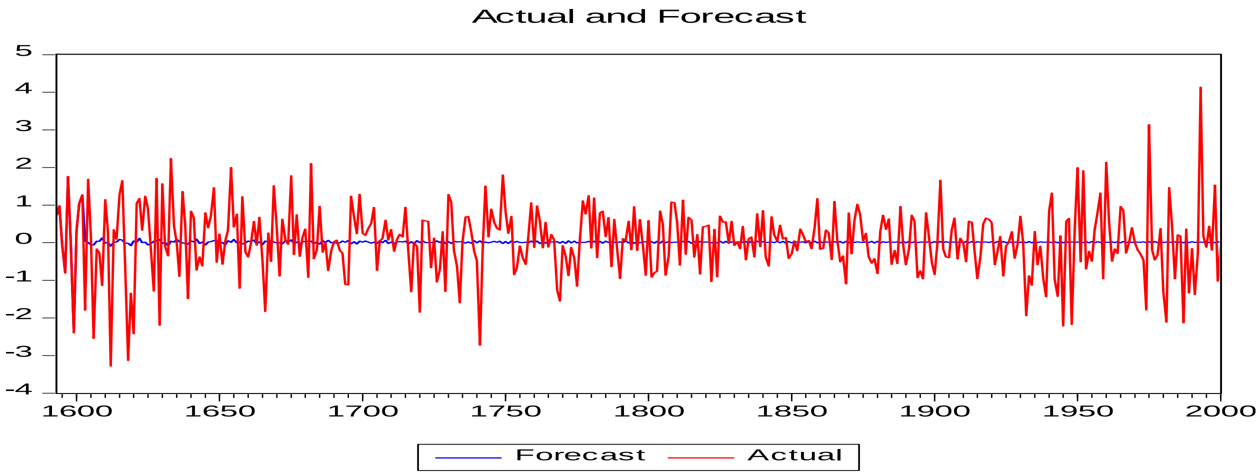

Figure 1: Comparison between actual and ARMA forecast series for the South African market.

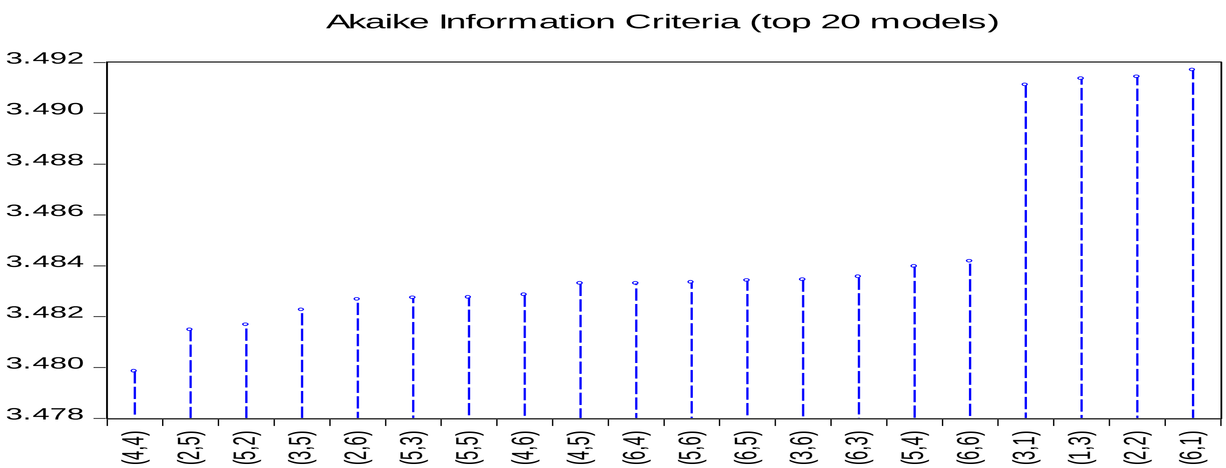

Figure 2: ARMA criteria graph for the South African dataset.

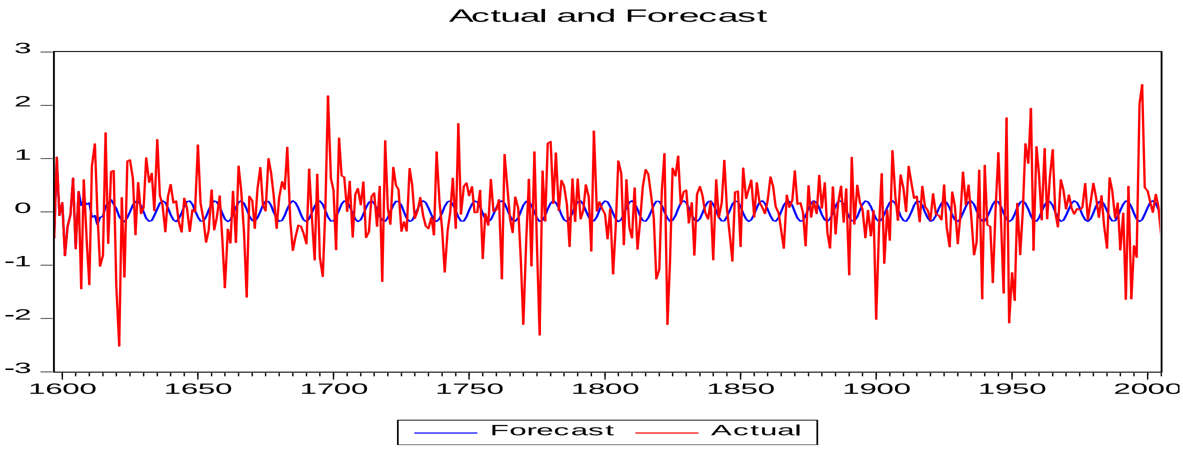

Figure 3: ARMA forecast comparison graph for US dataset.

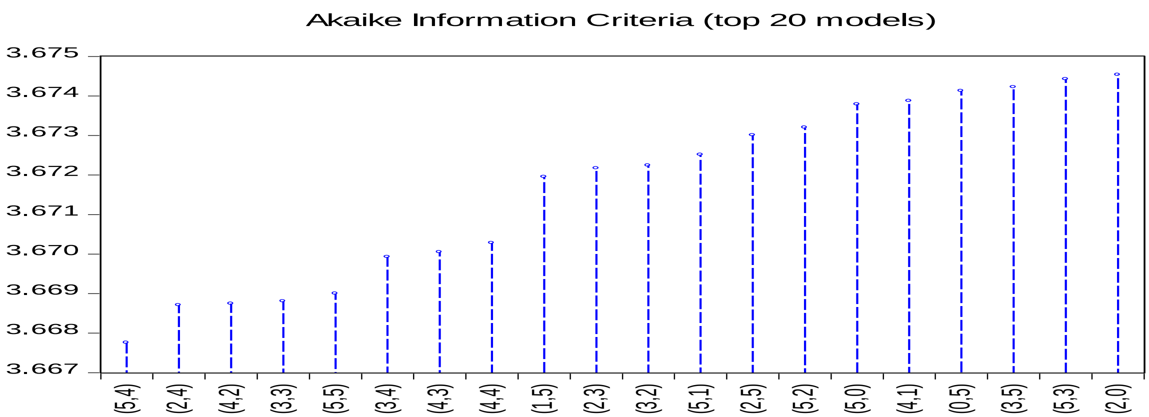

Figure 4: ARMA Criteria graph for US dataset.

3.2 <u>Artificial neural networks (ANNs)</u>

Artificial neural networks have been largely applied to various investment decision making and financial time-series forecasting ones. They have also become extremely popular in finance since financial services organizations have been the second largest sponsors of research for Neural Network applications. They were primarily developed in order to mimic the functionality of the human brain to a machine. The most attractive features of an ANN are enumerated below.

First, they are data-driven, self-adaptive models. Therefore, no a priori assumption about the statistical distribution of the data or any particular model specification is needed, during the model building process. Conversely, the network model can be adaptively formed taking into account the features presented from the data. Second, they are non-linear models, that assure accurate modeling of complex data patterns, compared to commonly used linear techniques such as ARIMA (Adhikari and Agrawal, 2013; Rozenberg *et al.*, 2012). Third, they are universal functional approximators, which enable them to approximate a large class of functions with high degree of accuracy (Khashei and Bijari, 2010). Furthermore, Neural Networks are less sensitive to error term assumptions and they can tolerate noise, chaotic components and heavy tails better than most other methodological approaches. Some additional advantages include larger fault tolerance, robustness and adaptability compared to expert systems due to the large number of interconnected processing elements that can be "trained" to learn new patterns (Kaastra and Boyd, 1996).

However, ANNs exhibit several drawbacks, from a statistical point of view. A commonly cited one refers to the fact that the model parameters are difficult, if not impossible to interpret. Therefore, ANNs are referred to as "black box" models and are initially built for pattern recognition and forecasting. Another one concerns the risk of overfitting or underfitting the data. *Overfitting* occurs when the constructed model is fairly complex and may fit irregular or unpredicted noise in the data. In this case, the model will be less reliable for out-of-sample forecasting. *Underfitting* occurs when a model is excessively simple to capture the underlying trend of the data and does not fit them well enough.

There are numerous types of ANNs in the forecasting literature. In this paper, two ANN Architectures are applied, namely, the Elman Network (EN) and Higher Order Neural Network (HONN) in order to forecast the daily market risk premium. Due to the randomness of the time series seeking in forecasting, a nonlinear nonparametric modelization approach is selected. Some other ANN Architectures have been tested but did not achieve to outperform the benchmark naive model.

3.2.1 <u>Elman Network (EN)</u>

A simple Recurrent neural network (SRNN) structure was conceived and first presented by Jeff Elman, in a paper entitled "Finding structure in time" (Elman, 1990).

Assume that a three-layer network with the addition of a set of "context units" is used. There are also connections from the middle (hidden) layer to these "context units" fixed with a value of one, namely bias nodes.

At each step, the input is propagated feed-forwardly, and then a learning rule is applied. The fixed back connections result in the "context units" and always keep a copy of the previous values of the hidden units (since they propagate over the connections before the learning rule is applied). Therefore the network is able to maintain a state that allows it performing a task beyond the power of a standard Multilayer Perceptron (MLP) network (Cruse, 2006).

The transfer sigmoidal function that processes the inputs from the hidden layer to the output layer is given by:

$$S(x) = \frac{1}{1+e^{x}} \tag{8}$$

and the linear function that has transformed the inputs linear:

$$F(x) = \sum_i x_i \tag{9}$$

Finally the Error function that needs to be minimized is:

$$E(u_j, w_j) = \frac{1}{T}\sum (Y_t - \hat{Y}_t(u_j, w_j))^2 \tag{10}$$

where Y_t is the actual value, $\hat{Y}_t$ is the target/predicted value, u_j is weight from the input to hidden layer, w_j is the weight from the hidden to the output layer.

3.2.2 <u>Higher Order neural networks (HONN)</u>

Higher Order neural networks were primarily presented by Giles and Maxwell (1987). Although their financial applications remain quite limited, HONNs have been largely applied in classification and prediction, nonlinear simulation and pattern recognition in computer science and engineering.

More analytically, HONNs use joint activation functions which reduce the need to establish connections among inputs while training. This diminishes the number of free weights, leading to faster training than even simple feedforward networks. In addition, the reduction of free weights demonstrates that the overfitting and local optima problems that affect the results, can be avoided (Kamruzzaman *et al.*, 2006). HONNs are alternatively referred as "open box" models, where each neuron maps variables to a function through weights/coefficients without the use of hidden layers (Shawash, 2012). The non existence of hidden layers in this kind of network architecture allows the application of easier training methods, leads to faster convergence, reduces the network size, allows for more accurate curve fitting, as opposed to other NN architectures (Zhang *et al*., 2002). As in the case of Elman network, HONN uses transfer sigmoidal function, Error function (similar to Mean Square Error minimization) and linear function. This has been set in order to facilitate direct comparison between these NN architectures.

3.3 Training the Neural Network

The network training refers to the adjustment of its weights in a way that it maps the output value of the training dataset to the output or predicted value. The training of the network starts with randomly selected weights and continues by applying the backpropagation of errors algorithm (Beale *et al.*, 2013; Alexandridis and Zapranis, 2014). This algorithm searches those weights that minimize the Error function (10). It is possible that the network learns the training data exactly, or alternatively referred as "*overfitting*", taking into account the number of nodes in the hidden layer. Therefore, the network training has to stop early and this is accomplished by splitting the dataset into three subsets.

Initially, the training set serves to optimize the model and the backpropagation learning algorithm determines optimal weights from the initial random weights. Then, the test set stops the training set from overfitting problems ("*early stopping*"). In particular, the optimization of the training set stops when the training set arrives at the maximum positive return. Training and test sets are accounted as insample subset and are divided in such a way to avoid overfitting problems and assure that data patterns will be captured by the network model. Finally, the validation set simulates future values of the time series under study.

Both neural networks (EN and HONN) are trained 80 times. The best 10 performing networks are selected for the out-of-sample forecasting task. The selected ones have the lowest insample error. Furthermore, each networks' forecasts differ depending on the different architectures being tested and their initial random set of weights. Therefore, a simple average of the committee of these 10 NNs is presented in such a way as to eliminate possible outlier network and avoid the problem of local optima that might have arisen during the training process.

Dataset	Trading Dates	From	To
Total set	2000	02/01/2007	31/12/2014
Training set (insample)	1200	02/01/2007	17/10/2011
Test set (insample)	400	18/10/2011	24/05/2013
Validation set (out-of-sample)	400	27/05/2013	31/12/2014

Table 1: Neural Networks' training dataset for Johannesburg Stock Exchange.

Dataset	Trading Dates	From	To
Total set	2005	03/01/2007	31/12/2014
Training set (insample)	1203	03/01/2007	20/10/2011
Test set (insample)	401	21/10/2011	30/05/2013
Validation set (out-of-sample)	401	31/05/2013	31/12/2014

Table 2: Neural Networks' training dataset for US Stock Exchange.

4.1 Data

This section aims to provide a detailed description of the data used in this paper and the transformations that have been made. In particular, the market risk premium for both financial markets (South Africa and US) is defined via the historical risk premium approach, as follows:

$$MRP = R_m - R_f \tag{11}$$

where:

- MRP is the market risk premium
- R_m is the market rate of return
- R_f is the risk-free rate return

There are various ways in calculating and/or estimating the risk premium depending on the research question a researcher seeks to answer.

This approach is a backward-looking one that predicts the future values of the market risk premium depending on past information. It is a quite appealing framework for practical forecasting situations where data is abundant or easily available, even though the theoretical model of the underlying relationship appears to be unknown.

Another well known approach is the expected one. This approach is a forward-looking one and is based on agents future expectations about moves in the market risk premium. It is a well esteemed approach and largely applied by the majority of theoretical asset pricing models. In this type of approach lots of macroeconomists and financial economists impose the no negativity constraint to the estimated market/equity risk premium and estimate it by a historically long annualized time-series of data. Various other researchers collect survey data on the expected risk premium from financial market practitioners.

In addition, the required risk premium approach is the return of the portfolio over the risk-free rate required by an investor. This approach is largely used in Corporate Finance and Portfolio Management applications.

A short-term T-Bill yield is chosen as proxy for the risk-free rate return, since the paper is oriented towards short-term forecasting and the total dataset consists of eight years. The choice of a short-time horizon, as opposed to long time series that is usually used in order to calculate and/or estimate the market risk premium, has been done intentionally. The methodology of the calculation of the market return changes periodically and longer time horizons of data do not facilitate direct comparisons among the combined datasets. Moreover, longer datasets contain structural breaks as well as data for emerging markets are difficult and expensive to acquire for relatively long time horizons.

All data have been sourced from DataStream for the period from January 2007 through to December 2014. Daily closing prices (adjusted for dividends and splits) of the FTSE/JSE All Shares Index and the S&P 500 Composite Index, are chosen as proxies for the market rate return.

More precisely, the chosen proxies for the risk-free rate return for South Africa and US are: South Africa Treasury Bill 91 Days (Tender rates), Yield-to Redemption and US T-Bill Secondary Market 3 Months, Middle Rate.

The figures below, show the calculated daily historical market risk premium time series for South African and US financial markets:

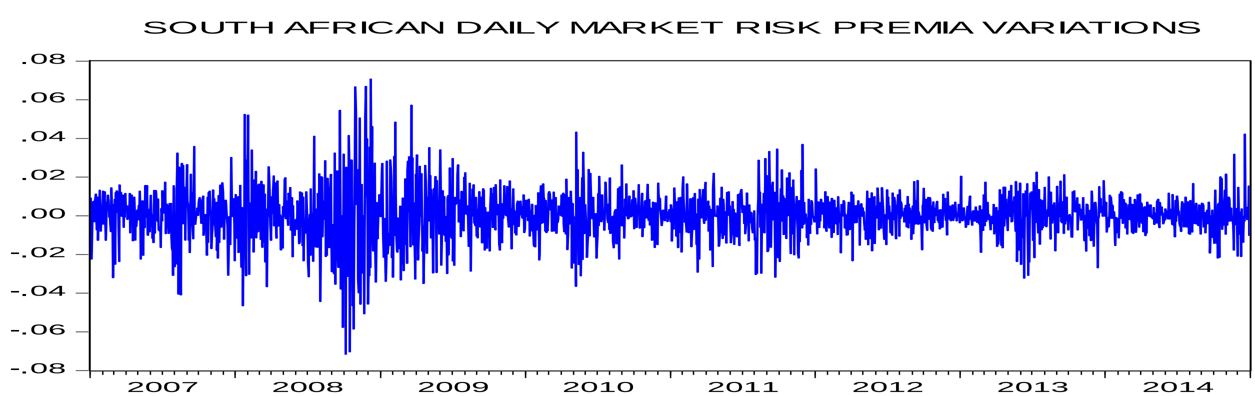

Figure 5: South African Market Risk Premium time series during the total sample period.

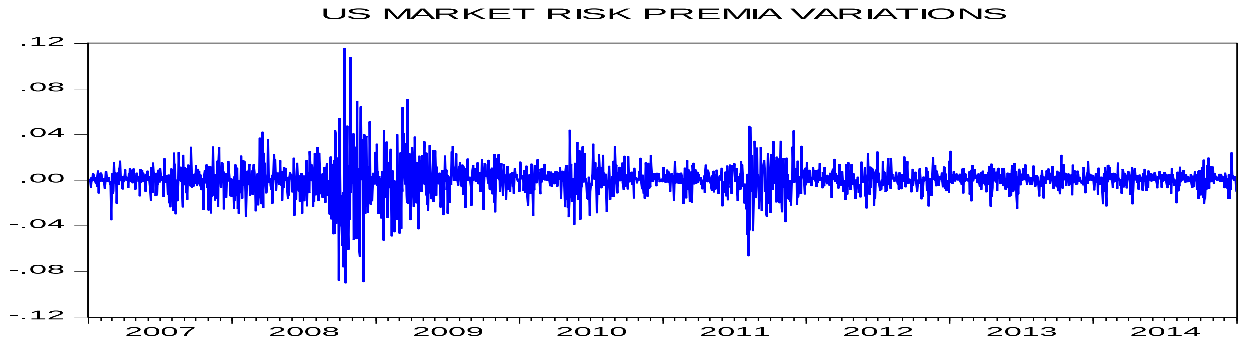

Figure 6: US Market Risk Premium time series during the total sample period.

The holidays and non-trading days are removed from the sample time-series. They contain high kurtosis, slight skewness and are non-normal. The Jarque-Bera test confirms it at 99 percent confidence interval, as presented by the summary statistics below:

DESCRIPTIVE STATISTICS		
	South Africa	***US***
Annualized mean	4.646628	6.095124
Annualized standard deviation	20.7094861	22.3605937
Median	0.051919	0.078778
Maximum	6.833477	10.95723
Minimum	-7.426423	-9.460417
Skewness	-0.121395	-0.312185
Kurtosis	6.639794	12.30121
Jarque-Bera	1108.921	7238.235
Probability	0.0000	0.0000
Number of observations	2000	2005

The daily closing prices, adjusted for dividends and splits, for each financial market are non-stationary and are transformed to daily series of rate returns in order to assure the stationary properties of these time series. Therefore, taking into account the Price level, $P_1, P_2, ..., P_t$, the return at time t is calculated as:

$$R_t = (\frac{P_t}{P_{t-1}}) - 1 \tag{12}$$

The selection of inputs (explanatory variables) inserted to the network has been done via feature/attribute selection. This method is used in order to apply measures to quantify the relevance of variables hidden in a large data set with respect to a given class or concept description. More precisely, feature selection serves as a filter muting out features that are not useful in addition to the existing ones. Such measures include information gain, the Gini index, uncertainty and correlation coefficients (Thawornwong and Enke, 2004). In this study, correlation coefficients are used in order to determine the necessary inputs that were included in the network. Therefore, the resulting variables having high correlation with the target/output time-series are chosen as the relevant input variables provided to the neural network models.

The number of lags has been selected after testing Autoregressive Models with different lags in an appropriate statistical software and selecting the best performing model with the lowest information criteria. The 9 and 10 main sector indices, respectively as well as the 91 day T-Bill redemption yield and the 3-month Treasury Bill are chosen as inputs to the neural networks since they have a clear explanatory power and correlation with the Market Risk Premium. The sector indices include the companies listed in the benchmark stock index. Therefore, sector indices are deemed as the most appropriate for the universal set of inputs.

In total, 10 and 11 inputs were inserted to the networks. Table 3 and Table 4 below show all the inputs to neural networks as well as the lags with which they were inserted. The final inputs are also normalized. This is intended for faster approaching to global minima at error surface and can also assure faster training.

The South African (SA) Sector Indices have split the All Share Index constituents according to their SA Sector classification. This classification is derived from the Industry Classification Benchmark (ICB). The US Sector Indices classification is based on GICS.

Number	Explanatory Variables	Lag*
1	FTSE/JSE Financials Total Return (TR)	2
2	FTSE/JSE Basic Materials TR	3
3	FTSE/JSE Consumption Goods TR	4
4	FTSE/JSE Consumption Services TR	3
5	FTSE/JSE Health Care TR	1
6	FTSE/JSE Industrials TR	2
7	FTSE/JSE Oil and Gas TR	3
8	FTSE/JSE Technology TR	1
9	FTSE/JSE Telecommunications TR	5
10	T-Bill 91 days (Tender Rates) - Redemption Yield	1

*Lag 1 means that today's return is used to forecast tomorrow's one.

Table 3: Explanatory Variables for Johannesburg Stock Exchange.

Number	Explanatory Variables	Lag
1	S&P 500 Consumer Discretionary TR	2
2	S&P 500 Consumer Staples TR	2
3	S&P 500 Energy TR	2
4	S&P 500 Financials TR	3
5	S&P 500 Health Care TR	3
6	S&P 500 Industrials TR	1
7	S&P 500 Information Technology TR	2
8	S&P 500 Materials TR	1
9	S&P 500 Telecommunication Services TR	2
10	S&P 500 Utilities TR	2
11	US T-Bill Secondary Market 3 months - Middle Rate	4

Table 4: Explanatory Variables for US Stock Exchange.

5. Empirical Results

The tested models are evaluated on their forecasting performance based on commonly used statistical metrics, namely, Mean Square Error (MSE), Mean Absolute Error (MAE), Directional Symmetry (DS), Correct Up Trend (CU) and Correct Down Trend (CD).

MSE and MAE measure the deviation between actual and forecast values. A small value of these measures indicates higher accuracy in forecasting. DS measures the correct direction of change predictions. CU measures the correct positive direction of change predictions and CD the correct negative direction of change predictions. The last three metrics are expressed in percentages.

Both neural network architectures (EN and HONN) were trained with 10 and 11 inputs each for South African and US financial markets respectively, one hidden layer and one output in order to predict the market risk premium for each market. The number of hidden nodes varied from 3~8 along the training of the networks and the HONN architecture was also tested for 2~4 orders. The neural networks were trained 80 times and 10 networks were finally chosen based on their insample performance. The number of hidden nodes of the 10 selected networks varied. In the South African case, the selected networks had 3, 5 and 7 hidden nodes, whereas in US, the hidden nodes were 4 and 6. In the case of HONN all selected networks were of order 4 for both markets. The tables 5 - 8 below show the insample and out-of-sample performance of the tested models compared to the benchmark model. The presented performance measures for EN and HONN models represent a combined forecast (simple average) of the 10 best performing networks out of the total 80 trained networks.

As shown in the tables 5-8 below, the neural network models achieve to provide superior results against the naive benchmark and the other tested models. It is clear that the best performer for both markets in both insample and out-of-sample results is the Elman network. The performance metrics of this network confirm its ability to generalize the result. I draw this conclusion from the fact that the directional accuracy performance metrics are within the same range of values in both the insample and the out-of-sample horizon periods.

	NAIVE	SES*	EWMA	ARMA(4,4)	EN	HONN
MSE	3,746	2,198	2,199	3,946	1,174	0,942
MAE	1,398	1,034	1,034	1,986	0,805	0,639
DS	43,683	51,226	51,228	48,774	56,254	56,505
CU	42,617	50,06	50,061	42,257	46,579	48,019
CD	44,914	52,576	52,576	44,650	50,594	49,67

*SES is Single Exponential Smoothing with decay parameter $a = 0,2$

Table 5: Insample statistical performance for the South African financial market.

	NAIVE	SES	EWMA	ARMA(4,4)	EN	HONN
MSE	1,659	0,811	0,887	0,803	0,322	0,355
MAE	0,966	0,901	0,715	0,664	0,46	0,451
DS	48,995	48,744	52,764	50,754	71,357	69,598
CU	45,714	45,714	99,048	50,952	68,571	67,619
CD	52,66	49,468	1,064	50,532	74,468	71,809

Table 6: Out-of-Sample statistical performance for the South African financial market.

	NAIVE	SES	EWMA	ARMA(5,4)	EN	HONN
MSE	5,189	2,704	2,739	4,556	1,174	2,206
MAE	1,53	1,05	1,058	1,423	0,805	0,99
DS	50	50,69	50,761	48,341	60,163	53,74
CU	47,429	49,371	48,854	48,229	57,829	51,657
CD	53,203	52,368	50,479	51,978	63,092	56,128

Table 7: Insample statistical performance for the US financial market.

	NAIVE	SES	EWMA	ARMA(5,4)	EN	HONN
MSE	0,995	0,559	0,563	0,505	0,516	0,531
MAE	0,764	0,55	0,56	0,532	0,539	0,553
DS	48,872	48,622	39,348	48,12	58,897	54,637
CU	44,589	47,619	39,827	48,485	60,173	53,680
CD	54,767	50	38,69	47,619	57,143	55,952

Table 8: Out-of-Sample statistical performance for the US financial market.

More analytically, the Higher Order neural network achieved to beat the naive benchmark model in both markets, whereas it provided inferior results compared to Elman network. The univariate ARMA, SES and EWMA models did not achieve to beat the naive benchmark in lots of instances.

The out-of-sample performance of the Elman network in the South African market exhibits Directional Symmetry of 71.357% whereas Higher Order neural network is of 69.598%. Correct Up Trend is 68.571% for EN and 67.619% for HONN. Correct Down Trend is 74.468% for EN and 71.809% for HONN.

In the US case, DS is 58.897% for EN and 54.637% for HONN. Correct Up Trend is 60.173% for EN and 53.68% for HONN. Correct Down Trend is 57.143% for EN and 55.952% for HONN.

The Directional Accuracy, Correct Up and Correct Down Trends also indicate the economic gains from applying neural networks in order to forecast the market risk premium in short-time horizons, where neural networks appear to outperform autoregressive moving average (ARMA) models.

In essence, modeling the market risk premium requires more than one variables in order to have a successful forecasting experiment. The results are quite promising in the case of South Africa compared to the ones in US. Taking into consideration that most of the linear equilibrium asset pricing models achieve around 60 percent forecasting accuracy, my results are within this range of forecasting performance, in particular in the case of US, whereas they exceed the expectations in the South African case.

A possible explanation behind the discrepancy of these markets might be their different standings in terms of market efficiency. Although this paper does not account for market efficiency testing in these markets, this can be one of the directions for future research work.

In economic terms, the examined datasets contain pre-crisis, crisis and recovery periods. By looking separately at the set periods (both insample and out-of-sample periods) and taking into account the self-adaptive properties of neural network models, I conclude that the repetitive training procedure simulates in a sort the search process that an investor could have achieved in real time to account for the possible structural changes (Qi, 1999).

Taking into consideration the above statistical performance of Neural Networks, the Diebold Mariano statistic for predictive accuracy is computed for both MSE and MAE loss functions. More information regarding the Diebold-Mariano statistic is presented in the Appendix.

The results of the Diebold-Mariano statistic, comparing all the other methods with the Naive model are summarized in the Tables 9 and 10 below:

South Africa	ARMA	SES	EWMA	EN	HONN
S_{MSE}	-9.3116	-7.1775	-6.6309	-9.1422	-9.9139
S_{MAE}	-12.0803	-7.7963	-6.6027	-11.4381	-12.8846

Table 9: Diebold Mariano statistics results for South African market

US	ARMA	SES	EWMA	EN	HONN
S_{MSE}	-6.2545	-4.0211	-6.4073	-6.0953	-5.7421
S_{MAE}	-7.9645	-2.9518	-8.0004	-7.518	-6.7673

Table 10: Diebold Mariano statistics results for US market

From the Tables 9 and 10 above, I note that the null hypothesis of equal predictive accuracy is rejected for all comparisons and for both loss functions at five percent confidence interval since the test results $|s_{MSE}| \succ 1.96$ and $|s_{MAE}| \succ 1.96$. Moreover, a negative realization of the Diebold-Mariano test statistic indicates that the first forecast is more accurate than the second forecast. The lower the negative value, the more accurate are the first forecasts.

In the Tables above, interesting patterns are apparent. According to the South African case, Neural Networks exhibit a clear superior performance compared to the other simpler univariate methods. However, in the case of US, results are mixed as both univariate (ARMA and EWMA) as well as Neural Networks exhibit high negative Diebold Mariano statistic.

After assessing the economic value of the forecasts I came up with the following results:

South Africa	Actual	Naïve	SES	ARMA	EWMA	EN	HONN
Mean	3,27%	3,59%	4,17%	3,23%	-26,37%	2,13%	6,76%
St. Dev.	89%	89,21%	25,71%	93,17%	5,63%	100,29%	63,91%

Table 11: Assessing the Economic Value of the Forecasts for the South African market

US	Actual	Naïve	ARMA	SES	EWMA	EN	HONN
Mean	6,56%	6,45%	1,26%	-29,6%	6,94%	-1,74%	-0,95%
St. Dev.	70,84%	71,02%	8,86%	21,16%	26,46%	10%	15,62%

Table 12: Assessing the Economic Value of the Forecasts for the US market

Based on the above descriptive statistics of the forecasts, it seems that the SES and HONN outperform the actual data and Naïve model and yield less volatility. Therefore, in terms of economic value, the SES and HONN models are selected for the South African market. However, compared to the actual data both models show signs of over prediction. ARMA is also close to the actual arithmetic mean of the realized market risk premium, whereas its volatility exceeds the actual one.

In the case of US market, EWMA is clearly the only model that achieves to outperform the Naïve one, as well as the actual data. In addition, its return is close to the one yielded by the actual data as well as its volatility is significantly lower.

6. Concluding Remarks

This paper has developed, applied and compared two machine learning models based on Artificial neural network Architectures, in order to forecast the market risk premium for the South African and the US equity markets, using the main sector indices and the 3-month T-bill yield. Univariate Single Exponential Smoothing, Exponential Weighted Moving Average and ARMA models were also tested against the benchmark naive model.

Results confirm the superiority of Artificial neural networks over simpler univariate forecasting models, as well as over the benchmark naive model. The Elman network outperforms all the other models in both insample and out-of-sample periods as well as between the tested markets.

The proposed forecasting model succeeds in capturing patterns in the data that better forecast the market risk premium across the South African and US financial markets. The forecasting accuracy in terms of sign accuracy are exceptional in the South African case. Directional Accuracy measures can thus be used as an indicator for trading and investment purposes.

The limitations of this methodological approach stem from the methodology itself. Unlike most of the traditional model-based forecasting techniques, Artificial neural networks are a class of data-driven self-adaptive and non-linear methods that do not require specific assumptions on the underlying data generating process. The results of the forecasting model are sensitive to the inputs selected and the time period chosen.

Some propositions for future work might account for the inclusion of some economic as well as financial variables in the forecasting experiment of the market risk premium. The models can also be tested in monthly frequencies and appropriate investment strategies can be introduced and tested against a naive buy-and-hold strategy.

References

1. Y. Abu-Mostafa and A. Atiya (1996), 'Introduction to Financial Forecasting', *Applied Intelligence* 6, 205 — 213.

2. R. Adhikari and R. K. Agrawal (2013), *An Introductory Study on Time Series Modeling and Forecasting*, LAP LAMBERT Academic Publishing.

3. A. K. Alexandridis. and A. D. Zapranis (2014), *Wavelet Neural Networks: With Applications in Financial Engineering, Chaos and Classification,* 1st ed., Wiley.

4. M. H. Beale, M. T. Hagan and H. B. Demuth (2013), *Neural Network Toolbox 8 User's Guide*, The MathWorks, Inc., Natick.

5. S. Bekiros and D. Georgoutsos (2008), 'Direction-of-Change Forecasting using a Volatility-Based recurrent neural network', *Journal of Forecasting*, John Wiley & Sons, Ltd., 27(5), 407 — 417.

6. T. Chenoweth and Z. Obradovic (1996), 'A multi-component nonlinear prediction system for the S&P 500 Index', *Neurocomputing* 10, 275 — 290.

7. H. Cruse (2006), *Neural Networks as Cybernetic Systems*, Second and Revised Edition.

8. G. Cybenko (1989), 'Approximation by superpositions of a sigmoidal function', *Math. Control Signal Systems* 2, 303 — 314.

9. A. Damodaran (2013), *Equity Risk Premiums (ERP): Determinants, Estimation and Implications – The 2013 Edition* Stern School of Business.

10. C. Dunis, A. Karathanasopoulos, G. Sermpinis and J. Laws (2014), 'Modeling and Trading the Greek Stock Market with Gene Expression and Generic Programming Algorithms', *Journal of Forecasting*, 33(8), 596 — 610.

11. J. L. Elman (1990), “Finding Structure in Time”, *Cognitive Science*, 14, 179 — 211.

12. E. L. de Faria, M. P. Albuquerque, J. L. Gonzalez, J.T.P. Cavalcante and M.P. Albuquerque (2009), 'Predicting the Brazilian Stock Market through Neural Networks and Adaptive Exponential Smoothing Methods', *Expert Systems with Applications*, 12506 — 09.

13. C. I. Giles and T. Maxwell (1987), ‘Learning, Invariance and Generalization in High-Order Neural Networks’, *Applied Optics*, 26(23), 4972 — 4978.

14. C. Hæke and C. Helmenstein (1996), 'Neural networks in the capital markets: An application to index forecasting', *Computational Economics*, 9(1), 37 — 50.

15. I. Kaastra and M. Boyd (1996), 'Designing a neural network for forecasting financial and economic time-series', *Neurocomputing*, 10, 215 — 236.

16. J. Kamruzzaman, R. Begg and R. Sarker (2006), *Artificial Neural Networks in Finance and Manufacturing*, Idea Group Inc, US.

17. Y. Kara, M. A. Boyacioglu and Ö. K. Baykan (2011), 'Predicting direction of stock price index movement using Artificial Neural Networks and Support Vector Machines: The sample of the Istanbul stock exchange', *Expert Systems with Applications* 38 (5), 5311 — 5319.

18. M. Khashei and M. Bijari (2010), 'An artificial neural network (p,d,q) model for time-series forecasting', *Expert Systems with Applications*, 37, 470 — 489.

19. K. M. Kiani and T. L. Kastens (2008), 'Testing Forecast Accuracy of Foreign Exchange Rates: Predictions from Feed Forward and Various Recurrent Neural Network Architectures', *Computational Economics*, 4, 383 — 406.

20. K. Kumar and J. D. Haynes (2003), 'Forecasting credit ratings using an ANN and statistical techniques', *International Journal of Business Studies*, 11(1), 91 — 108.

21. S. Mirmirani and H. C. Li (2004), 'Gold Price, Neural Networks ad Genetic Algorithm', Computational Economics, 23, 193 — 200.

22. L. Motiwalla and M. Wahab (2000), ' Predictable variation and profitable trading of US equities: A trading simulation using neural networks', *Computers and Operations Research* 27(11): 1111 — 1129.

23. C. J. Neely, D. E. Rapach, J. Tu and G. Zhou (2014), 'Forecasting the Equity Risk Premium: The Role of Technical Indicators', *Management Science*, 60(7), 1772 — 1791.

24. D. Olson and C. Mossman (2003), 'Neural Network forecasts of Canadian stock returns using accounting ratios', *International Journal of Forecasting* 19(3), 453 — 465.

25. M. Qi (1999), 'Nonlinear Predictability of Stock Returns Using Financial and Economic Variables', *Journal of Business & Economic Statistics*, 17 (4), 419 — 429.

26. D. E. Rapach, J. K. Strauss and G. Zhou (2010), 'Out-of-Sample Equity Premium Prediction: Combination Forecasts and Links to the Real Economy', *Review of Financial Studies*, 23(2), 821-862.

27. T. H. Roh (2007), 'Forecasting the volatility of stock price index', *Expert Systems with Applications* 33(4), 916 — 922.

28. G. Rozenberg, T. Bäck and J. Kok (2012), *Handbook of Natural Computing*, Springer Berlin Heidelberg, 461 — 477.

29. J. Shawash (2012), Generalized correlation higher order neural networks, neural network operation and Levenberg-Marquardt training on field programmable gate arrays, Doctoral thesis, UCL (University College London).

30. M. Spiegel (2008), 'Forecasting the Equity Premium: Where we stand today', *Review of Financial Studies,* 21(4), 1453 — 59.

31. S. Thawongnwong and D. Enke (2004), 'The adaptive selection of financial and economic variables for use with artificial neural networks', *Neurocomputing*, 56, 205 — 232.

32. H. Thinyne and J. Millin (2011), 'An Investigation into the Use of Intelligence Systems for currency Trading', *Computational Economics*, 37, 363 — 374.

33. R. Tsaih *et al.* (1998), 'Forecasting S&P 500 Stock Index Futures with a hybrid AI system', *Decision Support Systems*, 23, 161— 174.

34. S. Walczak (1999), 'Gaining competitive advantage for trading in emerging capital markets with neural networks', *Journal of Management Information Systems*, 16(2), 178 — 194.

35. I. Welch and A. Goyal (2006), 'A Comprehensive Look at the Empirical Performance of Equity Premium Prediction', *Review of Financial Studies* 21(4), 1510 – 31.

36. D. Witkowska and E. Marcinkiewicz (2005), 'Construction and evaluation of trading systems: Warsaw index futures', *International Advances in Economic Research* 11(1), 83-92.

37. M. Zhang, S. Xu and J. Fulcher (2002), ‘Neuron-adaptive higher order neural-network models for automated financial data modeling’, *IEEE Transactions on Neural Networks*, 13(1): 188 – 204.

38. G. Zhang, B. E. Patuwo and M. Y. Hu (1998) 'Forecasting with artificial neural networks: the state of the art', *International Journal of Forecasting*, 14, 35 — 62.

APPENDIX

A.1 Estimation Output and Model Selection for ARMA for the South African market risk premium

Dependent Variable: MRP_SA
Method: ARMA Maximum Likelihood (BFGS)
Sample: 1 1600
Included observations: 1600
Convergence achieved after 158 iterations
Coefficient covariance computed using outer product of gradients

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.014963	0.035024	0.427217	0.6693
AR(1)	-0.772943	0.013033	-59.30475	0.0000
AR(2)	0.203301	0.017596	11.55379	0.0000
AR(3)	-0.730878	0.016874	-43.31355	0.0000
AR(4)	-0.969394	0.012656	-76.59433	0.0000
MA(1)	0.797358	0.017835	44.70795	0.0000
MA(2)	-0.166009	0.024579	-6.753974	0.0000
MA(3)	0.723628	0.024036	30.10570	0.0000
MA(4)	0.940598	0.017912	52.51126	0.0000
SIGMASQ	1.875654	0.044842	41.82803	0.0000

R-squared	0.026815	Mean dependent var	0.014892
Adjusted R-squared	0.021313	S.D. dependent var	1.388718
S.E. of regression	1.373840	Akaike info criterion	3.479853
Sum squared resid	3004.797	Schwarz criterion	3.513430
Log likelihood	-2777.362	Hannan-Quinn criter.	3.492321
F-statistic	4.873887	Durbin-Watson stat	1.991484
Prob(F-statistic)	0.000002		

Inverted AR Roots	.57-.81i	.57+.81i	-.96-.29i	-.96+.29i
Inverted MA Roots	.55-.80i	.55+.80i	-.95-.28i	-.95+.28i

Table 9: ARMA(4,4) estimation output for the South African market risk premium time series.

Model Selection Criteria Table
Dependent Variable: MRP_SA
Sample: 1 1600
Included observations: 1600

Model	LogL	AIC*	BIC	HQ
(4,4)	-2777.362407	3.479853	3.513430	3.492321
(2,5)	-2779.665559	3.481480	3.511699	3.492701
(5,2)	-2779.826441	3.481681	3.511900	3.492902
(3,5)	-2779.295317	3.482266	3.515843	3.494734
(2,6)	-2779.623006	3.482675	3.516252	3.495143
(5,3)	-2779.672967	3.482738	3.516315	3.495205
(5,5)	-2777.690076	3.482759	3.523051	3.497720
(4,6)	-2777.771315	3.482861	3.523153	3.497821
(4,5)	-2779.128883	3.483307	3.520241	3.497021
(6,4)	-2778.129219	3.483307	3.523600	3.498268
(5,6)	-2777.167284	3.483355	3.527005	3.499563
(6,5)	-2777.220880	3.483422	3.527072	3.499629
(3,6)	-2779.248519	3.483456	3.520391	3.497170
(6,3)	-2779.338349	3.483568	3.520503	3.497283
(5,4)	-2779.668070	3.483980	3.520915	3.497694
(6,6)	-2776.824547	3.484175	3.531183	3.501630
(3,1)	-2790.385313	3.491118	3.511264	3.498598
(1,3)	-2790.582572	3.491364	3.511510	3.498844
(2,2)	-2790.638813	3.491434	3.511580	3.498915
(6,1)	-2787.854477	3.491703	3.521923	3.502924
(0,3)	-2792.163037	3.492089	3.508877	3.498322
(1,6)	-2788.179458	3.492109	3.522328	3.503330
(6,0)	-2789.199658	3.492134	3.518996	3.502108
(5,0)	-2790.215872	3.492155	3.515658	3.500882
(0,5)	-2790.340596	3.492310	3.515814	3.501038
(3,2)	-2790.342590	3.492313	3.515817	3.501040
(2,1)	-2792.348333	3.492320	3.509108	3.498554
(0,4)	-2791.349842	3.492322	3.512468	3.499802
(2,3)	-2790.357959	3.492332	3.515836	3.501059
(4,1)	-2790.385078	3.492366	3.515870	3.501093
(0,6)	-2789.532178	3.492550	3.519411	3.502523
(1,4)	-2790.582575	3.492612	3.516116	3.501340
(3,0)	-2792.609548	3.492646	3.509435	3.498880
(4,0)	-2791.627630	3.492669	3.512815	3.500149
(1,2)	-2792.844880	3.492940	3.509728	3.499174
(5,1)	-2789.864139	3.492964	3.519825	3.502938
(6,2)	-2787.928640	3.493044	3.526621	3.505512
(1,5)	-2790.108844	3.493269	3.520131	3.503243
(3,3)	-2790.204648	3.493389	3.520251	3.503363
(4,2)	-2790.384169	3.493613	3.520475	3.503587
(3,4)	-2789.486327	3.493741	3.523960	3.504961
(2,4)	-2790.572847	3.493849	3.520710	3.503823
(4,3)	-2790.180877	3.494608	3.524827	3.505829
(0,0)	-2798.706014	3.496512	3.503227	3.499005
(0,1)	-2797.870039	3.496717	3.506790	3.500457
(1,0)	-2797.911781	3.496769	3.506842	3.500509
(2,0)	-2797.283482	3.497233	3.510664	3.502220
(0,2)	-2797.439630	3.497428	3.510858	3.502415
(1,1)	-2797.796530	3.497873	3.511304	3.502860

A.2 Estimation Output and Model Selection Criteria for ARMA for the US market risk premium

Dependent Variable: MRP_US
Method: ARMA Maximum Likelihood (BFGS)
Sample: 1 1604
Included observations: 1604
Convergence achieved after 438 iterations
Coefficient covariance computed using outer product of gradients

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.014178	0.033017	0.429418	0.6677
AR(1)	0.036229	0.020520	1.765510	0.0777
AR(2)	0.660882	0.013294	49.71378	0.0000
AR(3)	-0.445786	0.004216	-105.7252	0.0000
AR(4)	-0.539741	0.000754	-715.9378	0.0000
AR(5)	-0.163524	0.013199	-12.38920	0.0000
MA(1)	-0.165128	0.122096	-1.352448	0.1764
MA(2)	-0.726791	0.177535	-4.093804	0.0000
MA(3)	0.579933	0.060518	9.582754	0.0000
MA(4)	0.530582	0.194720	2.724852	0.0065
SIGMASQ	2.254997	0.402219	5.606396	0.0000

R-squared	0.038208	Mean dependent var	0.014156
Adjusted R-squared	0.032175	S.D. dependent var	1.531679
S.E. of regression	1.506837	Akaike info criterion	3.667761
Sum squared resid	3619.270	Schwarz criterion	3.704639
Log likelihood	-2932.378	Hannan-Quinn criter.	3.681453
F-statistic	6.332378	Durbin-Watson stat	1.999753
Prob(F-statistic)	0.000000		

Inverted AR Roots	.79-.61i	.79+.61i	-.37-.26i	-.37+.26i
	-.82			
Inverted MA Roots	.79-.61i	.79+.61i	-.71+.16i	-.71-.16i

Table 10: ARMA(5,4) estimation output for the US market risk premium time series.

Model Selection Criteria Table
Dependent Variable: MRP_US
Sample: 1 1604
Included observations: 1604

Model	LogL	AIC*	BIC	HQ
(5,4)	-2932.379701	3.667763	3.704641	3.681455
(2,4)	-2936.137477	3.668707	3.695528	3.678665
(4,2)	-2936.171220	3.668749	3.695570	3.678707
(3,3)	-2936.216730	3.668806	3.695626	3.678764
(5,5)	-2932.374800	3.669003	3.709234	3.683940
(3,4)	-2936.118256	3.669929	3.700102	3.681132
(4,3)	-2936.216058	3.670051	3.700224	3.681254
(4,4)	-2935.401174	3.670282	3.703808	3.682729
(1,5)	-2938.743385	3.671954	3.698775	3.681912
(2,3)	-2939.918990	3.672173	3.695641	3.680886
(3,2)	-2939.976976	3.672245	3.695713	3.680959
(5,1)	-2939.192948	3.672515	3.699335	3.682472
(2,5)	-2938.590039	3.673009	3.703183	3.684212
(5,2)	-2938.747931	3.673206	3.703379	3.684409
(5,0)	-2941.216397	3.673790	3.697258	3.682503
(4,1)	-2941.285685	3.673876	3.697344	3.682589
(0,5)	-2941.487729	3.674128	3.697596	3.682841
(3,5)	-2938.564282	3.674223	3.707749	3.686671
(5,3)	-2938.728060	3.674427	3.707953	3.686875
(2,0)	-2944.816942	3.674538	3.687949	3.679517
(1,4)	-2941.877679	3.674614	3.698082	3.683327
(4,5)	-2938.164516	3.674971	3.711850	3.688663
(3,0)	-2944.350976	3.675204	3.691967	3.681427
(2,1)	-2944.429977	3.675302	3.692065	3.681526
(0,2)	-2945.481096	3.675366	3.688776	3.680345
(0,3)	-2944.801788	3.675765	3.692528	3.681989
(1,2)	-2944.911403	3.675902	3.692665	3.682126
(2,2)	-2943.915731	3.675907	3.696023	3.683376
(4,0)	-2944.292713	3.676377	3.696493	3.683846
(3,1)	-2944.341561	3.676438	3.696554	3.683906
(1,1)	-2946.398220	3.676509	3.689919	3.681488
(0,4)	-2944.455511	3.676580	3.696696	3.684048
(1,3)	-2944.737551	3.676932	3.697047	3.684400
(0,1)	-2948.477969	3.677854	3.687912	3.681588
(1,0)	-2950.281008	3.680101	3.690159	3.683835
(0,0)	-2961.211525	3.692475	3.699181	3.694965

Table 11: Model Selection Criteria for ARMA(5,4) US market risk premium model.

B. Statistical Performance Metrics

$$MSE = \frac{1}{n}(X_i - \hat{X}_i)^2$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|X_i - \hat{X}_i\right|$$

$$DS = \frac{100}{n}\sum_{i=1}^{n} d_i, \qquad d_i = \begin{cases} 1, & if \quad (X_i - X_{i-1})(\hat{X}_i - \hat{X}_{i-1}) \geq 0 \\ 0, & otherwise \end{cases}$$

$$CU = 100\frac{\sum_{i=1}^{n} d_i}{\sum_{i=1}^{n} t_i} \qquad d_i = \begin{cases} 1, & if \quad (X_i - \hat{X}_{i-1}) > 0, \quad (X_i - X_{i-1})(\hat{X}_i - \hat{X}_{i-1}) \geq 0 \\ 0, & otherwise \end{cases}$$

$$t_i = \begin{cases} 1, & if \quad (X_i - X_{i-1}) > 0 \\ 0, & otherwise \end{cases}$$

$$CD = 100\frac{\sum_{i=1}^{n} d_i}{\sum_{i=1}^{n} t_i} \qquad d_i = \begin{cases} 1, & if \quad (X_i - \hat{X}_{i-1}) < 0, \quad (X_i - X_{i-1})(\hat{X}_i - \hat{X}_{i-1}) \geq 0 \\ 0, & otherwise \end{cases}$$

$$t_i = \begin{cases} 1, & if \quad (X_i - X_{i-1}) < 0 \\ 0, & otherwise \end{cases}$$

Table 12: Statistical performance metrics that evaluate the forecasting accuracy of the tested models.

C. Diebold-Mariano statistic for predictive accuracy

The Diebold-Mariano statistic tests the null hypothesis of equal predictive accuracy. If n is the sample size and $e_i^1, e_i^2 (i = 1, 2, ..., n)$ are the forecast errors of the two competing forecasts then the loss functions are estimated as:

$$L_1^{MSE}(e_i^1) = (e_i^1)^2, \quad L_2^{MSE}(e_i^2) = (e_i^2)^2$$

$$L_1^{MAE}(e_i^1) = \left|e_i^1\right|, \quad L_2^{MAE}(e_i^2) = \left|e_i^2\right|$$

The Diebold-Mariano statistic is based on the loss differentials:

$$d_i^{MSE} = L_1^{MSE}(e_i^1) - L_2^{MSE}(e_i^2)$$

$$d_i^{MAE} = L_1^{MAE}(e_i^1) - L_2^{MAE}(e_i^2)$$

The tested null hypothesis which is based on the s_{MSE} and the s_{MAE} are:

H_o: $E(d_i^{MSE}) = 0$ against the alternative H_1: $E(d_i^{MSE}) \neq 0$

H_o: $E(d_i^{MAE}) = 0$ against the alternative H_1: $E(d_i^{MAE}) \neq 0$

The Diebold-Mariano test statistic s is estimated as:

$$s = \frac{d_i}{\sqrt{\hat{V}(\bar{d}_i)}} -> N(0,1)$$

where:

$$V(d_i) = n^{-1}[\hat{\gamma}_0 + 2\sum_{k=1}^{n-1}\hat{\gamma}_k] \qquad \text{and} \qquad \gamma_k = n^{-1}\sum_{i=k+1}^{n}(d_i - \bar{d}_i)(d_{i-k} - \bar{d}_i)$$