

Nonlinearities In Stock Return Prediction: Evidence From South Africa

Kathleen Hodnett, PhD, University of the Western Cape, South Africa
Heng-Hsing Hsieh, PhD, CFA, University of the Western Cape, South Africa
Paul van Rensburg, PhD, University of Cape Town, South Africa

ABSTRACT

This research investigates the relationship between firm-specific style attributes and the cross-section of equity returns on the JSE Securities Exchange (JSE) over the period from 1 January 1997 to 31 December 2007. Both linear and nonlinear stock selection models are constructed based on the cross-section of equity returns with firm-specific attributes as model inputs. Both linear and nonlinear models identify book-value-to-price and cash flow-to-price as significant styles attributes that distinguish near-term future share returns on the JSE. The risk-adjusted performance of the nonlinear models is found to be comparable with that of linear models. In terms of artificial neural network modeling, the extended Kalman filter learning rule is found to outperform the traditional backpropagation approach. This finding is consistent with our prior findings on global stock selection.

Keywords: Artificial Neural Networks; Genetic Algorithms; Stock Selection Models; Risk-Adjusted Performance

INTRODUCTION

Modern portfolio theory of Markowitz (1952) and the separation theorem of Tobin (1958) describe the asset allocation decision of rational investors in an efficient capital market. The asset pricing relationship depicted by the capital asset pricing model (CAPM) of Sharpe (1964) and Lintner (1965), along with the efficient market hypothesis (EMH) of Fama (1970, 1991), form the joint hypothesis for tests of market efficiency. The hypothesis states that firm-specific information is reflected in the share price in an efficient capital market, and firm-specific risk is diversifiable in a large portfolio. As a result, the only relevant risk is market risk, measured by an asset's beta coefficient against the movements of the market portfolio.

Tests of the joint hypothesis reveal several anomalies relating to the use of firm-specific information to earn abnormal returns above the risk-adjusted returns suggested by the CAPM. The CAPM-related anomalies include the value effect of Basu (1977), the small firm effect of Banz (1981), the long-term price reversals of De Bondt and Thaler (1985, 1987) and the momentum effect of Jegadeesh and Titman (1993). Possible explanations for these anomalies include methodological bias in empirical research, misspecification of the CAPM pricing relationship and the argument of investor irrationality from the behavioural perspective. Further to this, the existence of nonlinearity in the pricing relationship based on firm-specific attributes is observed in studies conducted by Banz (1981), Fama and French (1992, 1993), Van Rensburg and Robertson (2004), Eakins, Stansell and Buck (2003), Cao, Leggio and Schniederjans (2005) and Cao, Parry and Leggio (2009).

This research explores the existence of nonlinearity on the JSE Securities Exchange (JSE) over the period 01 January 1997 to 31 December 2007 by investigating candidate factors that contribute to the variations in the cross-section of equity returns on the JSE. We extend our prior study on multifactor models to incorporate nonlinearity in the cross-section of JSE equity returns. By comparing the results of the linear and nonlinear modeling techniques, the relative abilities of various combinations of firm-specific style attributes in forecasting equity returns are evaluated.

LITERATURE REVIEW

Since the introduction of the Sharpe-Lintner (1964, 1965) capital asset pricing model, numerous studies testing the predictability of stock returns in both a linear single factor and multifactor framework have emerged. However over the last few decades, evidence of anomalies have cast doubt on the validity of the joint (EMH-CAPM) hypothesis. Tests conducted by Banz (1981) find evidence of a size effect where a negative relationship between firm size (as measured by market capitalization) and average stock returns exists for stocks listed on the New York Stock Exchange (NYSE), over the period from 1927 to 1975. Fama and French (1992) find evidence of a size effect and a value effect for stocks listed on the NYSE, American Stock Exchange (AMEX) and the over-the-counter NASDAQ stocks over the period from 1963 to 1990. The authors report that both size and book-to-market ratio capture the cross-sectional return variation better than other combinations of variables, while beta possesses almost no explanatory power over the examination period. De Bondt and Thaler (1985, 1987) find evidence of long-term price reversals, while Jegadeesh and Titman (1993, 2001) find evidence of a short-term momentum effect. In addition to U.S. studies, these anomalies are evident on international economies. The two main branches of studies that attempt to explain the CAPM-related anomalies using factor models include the use of factor mimicking portfolios to represent priced risks proposed by Fama and French (1993) and the characteristic approach of Daniel and Titman (1997). While Fama and French (1993) use the size and value factor mimicking portfolios to explain the residual returns of the CAPM, Daniel and Titman (1997) argue that it is the style attributes themselves that explain the cross-section of equity returns. In terms of South African stocks listed on the JSE Securities Exchange (JSE), Van Rensburg and Robertson (2003a) and Van Rensburg and Robertson (2004) find evidence that support the argument of Daniel and Titman (1997).

In a univariate study, Hodnett, Hsieh and van Rensburg (2012a) estimate and examine the consistency of the payoffs to firm-specific attributes for stock listed on the JSE over the period 1997 to 2007. 32 firm-specific attributes under examination are grouped into five categories, namely (1) fundamental values relative to share price, (2) solvency and liquidity, (3) fundamental growth, (4) size and return momentum and (5) consensus analyst forecast. The authors report that those firms possessing higher fundamental values relative to their share prices, higher dividend and earnings growth, lower market capitalization, higher short-term returns and those with higher earnings forecasts earn relatively higher returns in the subsequent period in a consistent manner. The only category that did not yield significant attributes is the solvency and liquidity category. Hodnett, Hsieh and van Rensburg (2012b) extend the study to a multivariate framework over the same examination period. A sixth category, namely, operating performance, is included in the research yielding 38 style attributes under examination. Employing the methodology of Haugen and Baker (1996), in conjunction with the stepwise variable selection technique suggested by Van Rensburg and Robertson (2004), the authors construct 2 expected return multifactor models to predict forward monthly returns, with the style inputs as the model inputs. The models are designed to select inputs that maximize the in-sample Grinold (1989) information ratio and the Qian and Hua (2003) information ratio, respectively. The models are then tested in the out-of-sample period from 2002 to 2007. The models are updated every 12 months, along with the identity of the variables in a rolling procedure. Results confirm that the value attributes are the most important contributors to equity return forecasting on the JSE over the examination period. The model that maximizes the in-sample Qian and Hua (2003) information ratio outperforms the model which maximizes the in-sample Grinold (1989) information ratio.

Over the years evidence has emerged suggesting that stock return predictability could possibly be explained by a model incorporating nonlinearity. For example, Banz (1981: 16) concludes, "*The size effect is not linear in the market proportion (or the log of the market proportion) but is most pronounced for the smallest firms in the sample*". Further motivation for an asset pricing model that accounts for nonlinear behavior also dates back to Fama and French (1992). Stehle, Bunke and Sommerfeld (1997) argues that Fama and French (1992) assume a nonlinear relationship between expected returns and the untransformed values of the independent variables by using the logarithm of size and the logarithm of book-to-market ratio as independent variables in their cross-sectional regression models. Further to this, Stehle *et al.* (1997) argues that Fama and French (1992) use this form of nonlinear relationship without providing any reason.

Tests conducted by Hung, Liang and Liu (1996) and Kanas and Yannopoulos (2001) support the use of artificial neural networks (ANN) in financial research. Hung, Liang and Liu (1996) conduct research on the Taiwan

Stock Exchange in order to determine the feasibility of integrating the portfolio management process and artificial neural networks. This approach includes blending the arbitrage pricing theory (APT), an artificial neural network (ANN) and a portfolio constructor. A fixed-architecture network is trained using the backpropagation algorithm. The results provide evidence that the integrated approach provides a more optimal solution than when the APT or ANN is used alone, by eliminating the shortcomings of each technique. Kanas and Yannopoulos (2001) construct both a linear and an ANN model in order to examine and compare (out-of-sample) monthly returns for the Dow Jones (DJ) and Financial Times (FT) indexes, using the lagged return series of dividends and trading volume as the explanatory variables. With regard to both indexes, the results reveal the superiority of the out-of-sample ANN forecasts. In addition to this, the ANN forecasts explain the forecasting errors of the linear models for both indexes. This is not the case with the linear model. The authors' findings provide support for the underlying nonlinear relation between stock returns and fundamental attributes.

In terms of the examination of firm-specific attributes (style attributes), Eakins, Stansell and Buck (1998) examine nonlinearities in financial data by examining a set of firm-specific attributes for firms invested in by institutional investors. The sample includes companies listed on the NYSE, AMEX or those trades through the national association of security dealers. The data used in the study details the institutional ownership of about 3000 firms for the period from 1988 to 1991. An ANN model is constructed which includes the lagged values of ten firm-specific variables as inputs, namely, net profit margin, operating profit margin, current ratio, total asset turnover, debt-to-asset ratio, return-on-assets, price-earnings ratio, market value, trading volume and percentage ownership as the model output. The model is trained via the gradient descent backpropagation learning algorithm. Motivation for an ANN model is due to results of the first phase of their study in which tests for linearity revealed that beta, firm size and trading liquidity are nonlinear in their relationship with institutional ownership. In terms of predictive power, the ANN model outperforms the tobit regression model over the examination period. In a follow-up study, Eakins and Stansell (2003), in an attempt to determine whether value stocks provide superior investment returns, construct an ANN model using the descriptors of value, that is, price-to-cash flow, price-to-book, dividend yield, earnings yield, sales and market capitalization as inputs in the neural network, while percentage total return is specified as the output. All stocks listed on COMPUSTAT over the period from 1975 to 1996, with a market capitalization value greater than a current value of US\$150 million are included in the study. The model is trained using the backpropagation training algorithm. The results provided evidence that the risk-adjusted returns of portfolios selected by the ANN model are greater than the results achieved by other forecasting models. Consistent with Fama and French (1992, 1993) value stocks provide higher returns with lower risk than can be obtained from the random walk process.

Cao, Leggio, Schniederjans (2005) examine firm-specific attributes in emerging economies over the period from 1999 to 2002. Daily closing prices, quarterly book value and common shares outstanding for 367 public corporations traded on the Shanghai Stock Exchange are downloaded. ANN models are constructed to predict stock price movement on the Shanghai Stock Exchange and compared to linear regression models. Four models are constructed, two linear and two ANN models. The first linear model constructed is a univariate CAPM where stock returns are regressed on beta. A multivariate model is a replication of the Fama and French (1993) three-factor model. The ANN models are the nonlinear counterparts of the univariate and multivariate models. The backpropagation algorithm is used to train the ANN models. The results indicate that, in terms of the linear models, the CAPM outperforms the three-factor model in stock return prediction. It is also found that the ANN models outperform their linear counterparts in terms of their predictive power. Cao, Parry and Leggio (2009) extend the methodology of Cao, Leggio, Schniederjans (2005) by comparing three linear models with three ANN models on the Shanghai Stock Exchange over two sub periods, namely, 1999 to 2002 and 2003 to 2008. The models include a linear and an ANN univariate time series model, a multivariate linear and nonlinear version of the CAPM (where stock returns are regressed on market returns), and the linear and nonlinear version of Fama and French (1993) three-factor model. The results once again confirm the forecasting superiority of the ANN models relative to the linear models. There is also no significant difference between the CAPM and the three-factor model, both in the linear and the nonlinear case.

Hodnett and Hsieh (2012) investigate the potential of ANNs in the stock selection process of global actively managed funds over the examination period from 2004 to 2009, using the Dow Jones Sector Titans as the research database. Two ANN models are constructed within the cascade-correlation architecture, with 17 style

attributes as candidate model inputs in the cross-section of global stock returns. The results support the use of artificial neural networks for financial forecasting in the stock selection process of actively managed funds. Particularly, the model trained via the embedded extended Kalman filter had greater strength in identifying future top performing stocks than the model trained via the back propagation learning rule.

Empirical studies in South Africa also provide ground for research into the nonlinearity of equity returns for stocks listed on the JSE Securities Exchange. For example, Fraser and Page (2000) study the momentum effect on the JSE over the period of 1977 to 1997. They find that although momentum shares with relatively higher prior returns earn higher future returns in general, the negative momentum shares earn higher returns than the shares placed in the fourth momentum quintile. Van Rensburg and Robertson (2004) find that the firms in the fifth (smallest) size and value quintiles earn higher returns than the firms in other quintiles over the period from July 1990 to June 2000. However, the average monthly returns earned by the respective quintiles do not appear to be linear. More specifically, the shares in the third size quintile actually earn lower returns than shares in the second size quintile. On the other hand, the average monthly returns are progressively distributed to the shares in the higher value quintile (that is, higher indications of value). If these nonlinearities are found to be apparent, of more relevance then is to determine whether these nonlinearities are robust over time, and how to capture and utilize them in forecasting equity returns and subsequent stock selection.

ARTIFICIAL NEURAL NETWORKS BACKGROUND

Artificial neural networks represent abstractions of biological neural networks and are designed to imitate the functioning of the human brain. ANNs are composed of processing elements operating in parallel, which have the ability to map any arbitrary nonlinear function. ANNs are composed of neurons, commonly referred to as ‘nodes’, ‘processing elements’ or ‘perceptrons’. A biological neuron, in its most basic form, is a data-storing cell, responsible for the cognitive processes of the brain, such as memory, problem-solving or the ability to learn. An interconnected set or group of these neurons is referred to as a neural network.

An artificial neuron, represented in Figure 1, is the most basic component of an artificial neural network. Inputs ($x_0 \dots x_n$) are fed into the neuron, weighted and summed and then passed through a nonlinear activation function to the output (y). The connection (synaptic) weights ($w_0 \dots w_n$) represent the interconnection between the units. The weights are modified to represent the respective connection strengths. x_0 represents the bias inputs, while, w_0 represents the bias weight. Nonlinearity is introduced into the network by way of an activation function. The sigmoid activation function (ϕ) is the most commonly used transfer function in neural networks. In linear regression the activation function is the identity function.

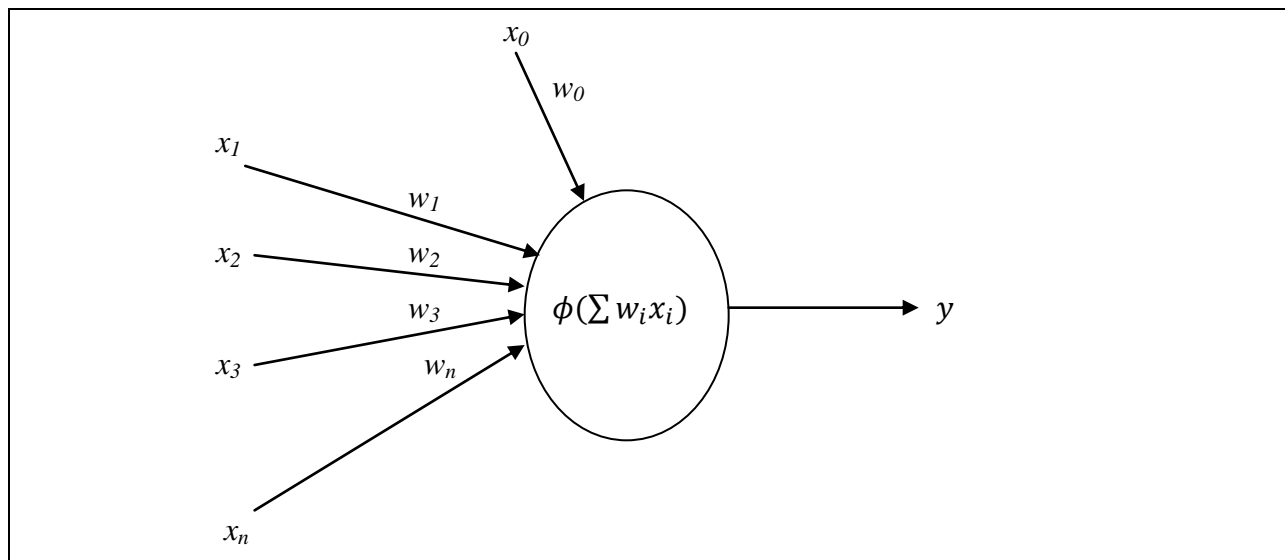


Figure 1: Structure of a Neuron

Thus, a group of artificial neurons, with weighted interconnection and structured in layers (consisting of an input layer, one or more hidden layers and an output layer) with parallel processing is referred to as an artificial neural network. Neural networks may differ depending on the network architecture and training mechanism. For example fixed-architecture feed-forward networks are traditional representations where information passes from the input layer neurons to the hidden layer neuron(s) and finally propagated to the output layer neuron(s), via interconnection weights. Neurons occupying a particular layer are in no way interconnected- information is fed forward in such a way that each input layer neuron is fully connected each hidden layer neuron, which in turn is connected to each neuron in the output layer. In some instances, in addition to the interconnections described, a direct connection may also exist between the neurons in the input layer and the neurons in the output layer.¹ This direct connection from input to output neurons captures any linear relationships in the data, while the incorporation of the hidden neurons captures nonlinearity. For example, Equation 1 (modified from Kanas and Yannopoulos (2001) and adapted from Hodnett and Hsieh (2012)) represents the expected returns determination process where a fixed-architecture ANN model with direct connections from input to output layer is used to estimate the expected return of share *i* in month *t*.

$$E(R_{it}) = a_{i,0,t} + \sum_{j=1}^J g_{i,j,t} F_{j,t-1} + \sum_{k=1}^K b_{i,k,t} f \left(d_{i,0,t} + \sum_{j=1}^J c_{i,j,k,t} F_{j,t-1} \right) \tag{1}$$

Where:

- $E(R_{it})$ = the expected return of share *i* for period *t*;
- $a_{i,0,t}$ = the bias term to the output layer for period *t*;
- $g_{i,j,t}$ = the direct payoff to input neuron $F_{j,t-1}$ from the output neuron (the equity returns) for period *t*;
- $F_{j,t-1}$ = the input neuron representing the transformed lagged value of attribute *j* in the input layer;
- $b_{i,k,t}$ = the factor payoff to the hidden neuron *k* from the output neuron for period *t* (it measures the sensitivity of the equity returns to movements in hidden neuron *k*);
- $d_{i,0,t}$ = the bias term to the hidden layer for period *t*;
- $c_{i,j,k,t}$ = the factor payoff to input neuron $F_{j,t-1}$ from the hidden neuron *k* for period *t* (it measures the sensitivity of hidden neuron *k* to movements in input $F_{j,t-1}$);
- $\sum_{j=1}^J g_{i,j,t} F_{j,t-1}$ = the linear function;
- $f \left(d_{i,0,t} + \sum_{j=1}^J c_{i,j,k,t} F_{j,t-1} \right)$ = the nonlinear activation function (the sigmoid function); and
- $d_{i,0,t} + \sum_{j=1}^J c_{i,j,k,t} F_{j,t-1}$ = the input signal within the nonlinear activation function.

The backpropagation learning algorithm is a gradient descent mechanism used to update the weights of the network, until the network error is minimized. The extended Kalman filter, based on state-space theory, is an alternative learning algorithm designed to update/adjust the network weights. Both mechanisms are illustrated in detail in Hodnett and Hsieh (2012) and discussed within the context of the cascade-correlation architecture presented in Figure 2.

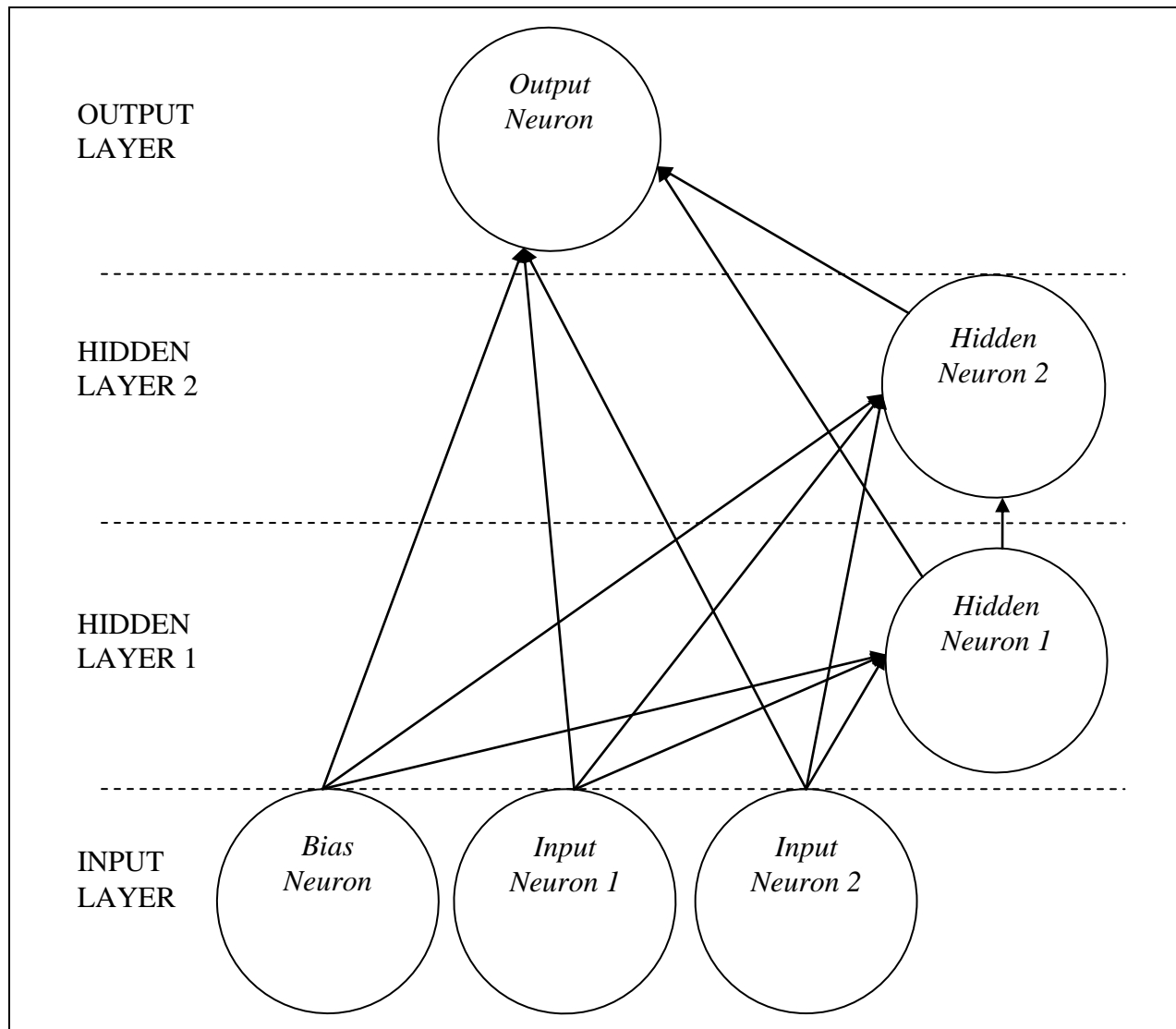


Figure 2 Cascade-Correlation Architecture: This topology-modifying architecture is an example of a feed-forward network. Activation is passed forward from the input layer to the first hidden neuron (first hidden layer). Activation then passes to the second hidden neuron (second hidden layer) and finally to the output neuron (output layer).

The structural design of the ANN architecture as well as the training rule/ mechanism influences the manner in which the network is able to learn. In addition to fixed-architecture networks, there exists the cascade-correlation architecture of Fahlman and Lebiere (1990/1991), presented in Figure 2. Cascade-correlation serves a two-fold purpose. Firstly, it establishes the cascade-architecture where hidden neurons are added one by one to the network. Once they have been added to the network, they do not change. Secondly, it is a feed-forward, supervised learning algorithm for training artificial neural networks. It is the job of the learning algorithm to install the hidden neurons. Many different networks are trained until the best one is selected. No *a priori* network architecture needs to be specified in advance (as is the case with fixed-architecture networks). The objective of constructing ANN models is to determine the existence of essential relationships in data. More complex patterns would thus require more hidden layers. With cascade-correlation networks, “*there can be any number of hidden nodes, depending on (and increasing with) the complexity of the pattern in the input data*” (Coats and Fant, 1993: 144). Initially there are no hidden nodes in the hidden layer. Hidden nodes are then added to the network one at a time. “*For each new hidden unit, we attempt to maximize the magnitude of the correlation between the new unit’s output and the residual signal we are trying to eliminate*” (Fahlman and Lebiere, 1990/1991: 3). Note in Figure 2 that each hidden neuron is fully

connected from both input neurons and pre-existing hidden neurons to all existing output neurons. The objective of this cascade architecture is thus “to “incrementally” capture the knowledge in the unique structure of hidden nodes and connection weights” (Coats and Fant, 1993: 145-147). This cascaded architecture also ensures that the neural network determines the actual number of hidden nodes and connection weights.

All ANN models in this research are trained within the cascade-correlation learning algorithm. This algorithm trains the weights using an embedded backpropagation or extended Kalman filter learning rule described in Hodnett and Hsieh (2012).

RESEARCH METHODOLOGY

In this study we analyze 4 expected return multifactor models constructed to predict forward monthly JSE sample stock returns, over the examination period 01 January 1997 to 31 December 2007, using firm-specific attributes as model inputs. These multifactor models constructed consist of 2 linear and 2 nonlinear models. Since our objective is to explore nonlinearities in the cross-section of JSE returns, we include the linear models developed in our prior study in order to draw a meaningful comparison with the forecasting and stock selection abilities of the nonlinear models. The linear models constructed in Hodnett, Hsieh and van Rensburg (2012b) is based on the methodology of Haugen and Baker (1996) and the stepwise variable selection procedure suggested by Van Rensburg and Robertson (2004). The first linear model, Grinold (IR), had the objective of selecting those variables which maximizes the in-sample Grinold (1989) information ratio, while the second linear model, QH (IR), had the objective of maximizing the in-sample Qian and Hua (2003) information ratio. These models were then evaluated over a specified out-of-sample period, namely 2002 to 2007. The models and style attributes were updated every 12 months based on training over the prior 60 months period using a stepwise variable selection procedure. In this paper, we construct 2 nonlinear models, which are the counterparts of the linear models, using artificial neural networks. These models are constructed within the cascade-correlation architecture of Fahlman and Lebiere (1990/1991), with different embedded learning rules, namely the traditional backpropagation learning rule and the extended Kalman filter learning rule. We utilize the same dataset set outlined in Hodnett, Hsieh and van Rensburg (2012b) which consists of closing stock prices, indexes and style attributes for 159 stocks comprising the FTSE/JSE All Share Index over the examination period from 01 January 1997 to 31 December 2007, downloaded from Datastream International. The 38 style attributes (extracted from Hodnett, Hsieh and van Rensburg (2012b)) are presented in the Appendix.²

As mentioned, the ANN models are constructed within the cascade-correlation architecture and trained using an embedded backpropagation and the extended Kalman filter learning rule, respectively. ANN_Backprop is trained via the backpropagation learning rule while ANN_Kalman is trained using the extended Kalman filter learning rule. The same evaluation criteria as the linear models, namely, the Grinold (1989) information ratio and the Qian and Hua (2003) information ratio are used to evaluate the performances of the ANN models. This allows direct comparisons to be made between the performances of the linear expected return factor models and their nonlinear counterparts presented in this paper. We also intend to draw a comparison with the results of Hodnett and Hsieh (2012) who found that the model trained via the extended Kalman filter rule outperformed the model trained via the backpropagation learning rule in the global equity market.

Consistent with the linear models, this study employs the moving-window procedure whereby 60 months data are trained to forecast the immediate 12-month returns following the training period (60-for-12). Under this design, the dataset spanning the period from 01 January 1997 to 31 December 2001 represents the first training dataset to forecast the 12 months returns in 2002. The second training window covers the dataset spanning the period from 01 January 1998 to 31 December 2002. This training window discards the first 12 months data (that is, the data in 1997) and incorporates new information that becomes available in 2002 to forecast the 12-month returns in 2003. This overlapping procedure is then repeated a further 4 times until all returns in the out-of-sample period is forecasted. Additionally, the out-of-sample period is also divided into two sub-periods from 2002 to 2004 and from 2005 to 2007.

ANN Model Design

Four areas of concern are addressed in the designing of the ANN models: selecting and specifying the data sets (train, test and validation sets), input variable selection, ANN construction and training; and ANN model evaluation.

(a) Train, Test and Validation Set Selection

When developing a model to project the expected return of each share over the in-sample period, the objective is to maximize the accuracy of the estimation over the in-sample period. As is with the factor selection procedure of the linear factor models, nonlinear models with differential style factors are constructed. Haugen and Baker (1996: 408) emphasize that *“in building an expected return factor model, we must estimate the tendency for stocks with differing exposures to different factors to produce differing returns”*. To satisfy this condition, the factor payoffs to style attributes have to be identical across all sample shares in any given month. This will ensure that shares with identical values of the style attributes employed in the model will produce identical expected returns from the model. The cross-sectional factor weights are estimated by training the MLP (multilayer perceptron) of the ANN models. The lagged values of the cross-sectional style attributes are used as the model inputs that correspond with the cross-sectional realized returns of the sample shares (that is, the “target” outputs). Three data sets are specified, namely, the training set (in-sample set), the test set (in-sample set) and the validation set (out-of-sample set). The purpose of the training set is to map the model parameters. The network weights are thus updated using the training set. The generalization ability of the learning process is assessed by the training set. The test set is indirectly used in building the model. The test set monitors model performance by observing how well the model interpolates. A technique is adopted whereby training set records are selected from a secondary working set at regular intervals. For all models, this procedure is such that 70% of the in-sample data set is used as the training set while 30% is selected as the test set at regular intervals. The validation set is completely independent of the model building process. The out-of-sample data, which represents the validation set, is manually set aside for verification of the network.

(b) Input Variable Selection

All 38 variables are selected as potential input variables in the neural network. The proper selection of inputs cannot only reduce the size and complexity of the network, but also lessen the chance of overfitting. Selecting those significant variables that act synergistically when combined with each other, is imperative in ANN modeling. In the past, different variable selection procedures have been employed in research entailing stock return prediction. For example, studies have employed stepwise and backward stepwise regression (Motiwalla and Wahab (2000)), principal component analysis, or selecting variables based on a set of *a priori* assumptions, to name a few. Some authors argue that although several available methods are applicable for linear modeling, these approaches cannot be simply extended to nonlinear modeling. For example, Mao and Billings (1999: 352) attribute the inapplicability of forward selection and backward elimination algorithms to the weight structure of an MLP neural network and explain *“these algorithms are in fact term selection and/or deletion methods rather than variable selection and/or deletion algorithms in the nonlinear system case”*.

In this paper we employ a genetic algorithm (Holland (1975) and Goldberg (1989)) to select the optimal subset of input variables for the neural network.³ This evolutionary search procedure has the capability of modeling complex systems with multiple solutions. Inspired by Charles Darwin’s principle of evolution and genetics, the laws of natural selection apply where only the fittest members of a group will survive. These fit members will genetically recombine with other fit members to produce offspring thereby ensuring that good characteristics will pass on to the next generation, while those less fit members will be eliminated. The standard GA starts with the simulation of a random population consisting of subsets of input variables referred to as “individuals” or “chromosomes” in the evolutionary process. Each member of the population is then evaluated and graded according to a fitness function. This ensures that only the proportion of the fittest individuals from the current (“parent”) population is extracted. Through the process of reproduction (crossover and mutation) individuals are selected for the next generation. This continues until a particular stopping criterion is achieved. This process is depicted in Figure 3.

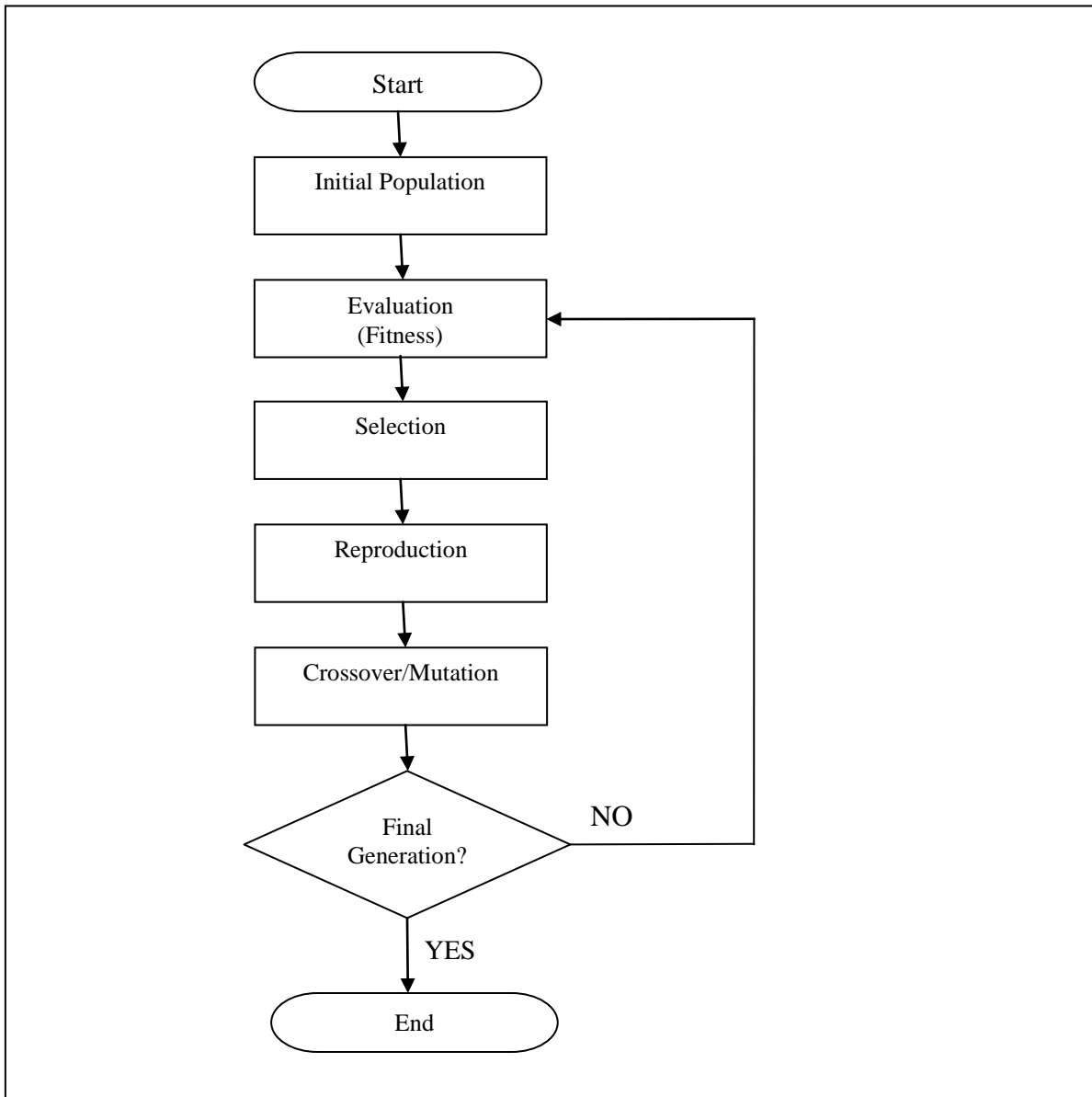


Figure 3 The Genetic Algorithm Cycle: The standard genetic algorithm begins with the simulation of a random population consisting of subsets of input variables (“individuals”). During successive iterations (“generations”) the initial individuals advance towards more fit individuals by reproduction among members of previous generation. Via genetic operators (selection, crossover and mutation) the next generation is created. Individuals are selected for the next generation, until a particular stopping criterion is attained.

The fitness function is imperative in genetic algorithms. With the exception of all elite individuals new individuals are either copied from a single fit parent (current individual) to the next generation or are the product of two fit parents that are recombined.⁴ The research computes probabilities of the first parent, Parent 1, according to rank, and the second parent, Parent 2, is selected randomly according to a uniform distribution. Computing probabilities according to fitness is generally problematic in generating an optimal solution and thus not implemented. In determining probabilities according to rank, individuals are first arranged according to relative fitness, and selected according to a linear distribution, favoring the higher ranked individuals. A linear rank bias (ranging between 1 and 2) of 1.4 is specified for both ANN models. The higher this value is, the more bias there is

towards choosing the fitter individuals as parents. With regard to ANN_Kalman, a necessary requirement for the extended Kalman filter rule is that the number of input variables presented to network for training is reduced. Thus a linear variable selection model is deemed inappropriate. Instead, the genetic algorithm will employ a neural network (secondary neural network) as the variable selection model to determine fitness.

Reproduction ensures that only the highest scoring parents will propagate their characteristics to the next generation. The most important genetic operators are the crossover and mutation operators. The crossover operator takes two parent individuals and combines their characteristics to produce new individuals (children/offspring). A cross-over probability of 0.7 is specified, specifying the likelihood that the next generation will be as a result of two parents from the previous generation combining and not as a result of replication from a single parent in the previous generation. A way of re-introducing lost variables back into the population is by specifying a mutation factor. The mutation operator is applied after the crossover operator and excludes any elite individuals. A mutation is a random addition or deletion of a characteristic in an individual, and is governed by a preset mutation factor. A mutation factor of 1 percent is specified. Another technique used in this research is the application of an “elitist factor”, which determines what fraction of the fittest members of the population will survive to be continued into the next generation via replication. An elitist factor of 0.05 is specified.

Jarvis and Goodacre (2007) suggest three ways in which the stopping criterion can be achieved. This includes setting the maximum number of iterations (known as generations) that the genetic algorithm must run for, specifying a maximum target outcome value for fitness, or a specifying a set number of generations for which the fitness value for the fittest individual remains constant. In specifying the convergence conditions for the genetic algorithm, the maximum iterations allowed for this procedure is 50 generations for ANN_Backprop and 30 generations for ANN_Kalman. Tolerance per generation is used to determine whether the average fitness has improved, while patience places a limit on the number of iterations allowed during variable selection. Under the condition where the average fitness of the population has not improved within the specified tolerance factor (0.001) for 50 iterations for ANN_Backprop or 30 for ANN_Kalman, the algorithm is halted.

(c) ANN Construction and Training

Cascade-correlation can be divided into two repetitive phases. The first phase involves the cascade-architecture, while the second phase involves the recruitment of a “candidate” neuron to the growing network architecture. In the first phase, the network starts with an input and an output layer and no hidden neurons (layers). All the connections (connecting input to output layer) are trained using an embedded backpropagation learning algorithm for ANN_Backprop and the extended Kalman filter rule for ANN_Kalman.⁵ This training continues until the error of the network no longer declines. If no significant error reduction has occurred within a specified number of epochs (training cycles), and there exists some residual error that needs to be further reduced, a hidden neuron (layer) may be added to the network. Once a hidden neuron is installed (activated), its input connections are frozen and the neuron is in no way altered/modified. The network is trained again and if no further reduction in the error occurs, another hidden neuron (in a new hidden layer) is installed, with this new neuron connected to the previous hidden neuron, resulting in the cascaded architecture (Refer to Figure 2). This process continues until the global minimum is established or a final architecture is obtained. After each hidden neuron is installed, training of the output weight resumes.

An important consideration is the determination of the hidden neuron (described in the first phase) to be installed in the hidden layer. Before a neuron is “recruited” for installation in the hidden layer, there may exist a number of “candidate” neurons, each possessing the potential to qualify as a hidden neuron. The second phase thus involves the analysis of a candidate neuron and can be divided into two parts. The first part is referred to as *input-to-hidden-layer training*, while the second part of this phase is referred to as *hidden-to-output-layer training*. *Input-to-hidden-layer training* starts with the generation of the candidate neurons. Each candidate neuron is connected with all input neurons and all existing hidden neurons by trainable input connections. At this stage, there exist no connection weights between the candidate neurons and the output neurons. The weights (on the input side) of the candidate neurons are trained using the backpropagation algorithm, with the objective of maximizing the correlation between the residual error of the network and the activation of the candidate neurons. Training halts when no further improvements to the correlation score results. The candidate unit with the maximum correlation is installed in the

network as a hidden neuron. The goal of the adjustment is to maximise S , the sum over all input units o of the magnitude of the correlation V , the candidate unit's value, and E_o , the residual output error observed at unit o . S is defined as follows (Fahlman and Lebiere (1990/1991)) :

$$S = \sum_o \left| \sum_p (V_p - \bar{V}) (E_{p,o} - \bar{E}_o) \right| \tag{2}$$

Where o is the network output at which the error is measured and p is the training pattern. The quantities \bar{V} and \bar{E}_o are the values of V and E_o averaged over all patterns.

Training is stopped if the correlation ceases to improve, or if the predefined number of cycles is exceeded. The final step of the second phase is the installment of the candidate neuron as a hidden neuron in the network. The newly trained weights are frozen and no further adjustments or modifications made to these weights. Freezing the input weights ensures that it will always be able to track the aspects of the network error that it was trained to track. The new hidden neuron now has trainable weights connecting it with the output neurons. These weights now undergo training (*hidden-to-output-layer training*). The addition of the hidden neurons continues until the learning process is complete.

(d) Evaluation of ANN Models

The overall accuracy of the ANN model predictions is determined by computing the Grinold (1989) Information ratio and the Qian and Hua (2003) information ratio. Both performance measures are derived from the information coefficient (IC) which is a measure of forecasting accuracy. The IC measures the monthly correlation between the realised returns and the forecasted returns. The Grinold (1989) information ratio provides an indication of a model's forecasting accuracy with adjustments to the model's breadth (the number of independent forecasts over the evaluation period).

$$\text{Grinold Information Ratio} = \frac{\sum_{t=1}^T \rho_t(R_{i,t}; E(R_{i,t})) \times \sqrt{N_t}}{T} \tag{3}$$

Where:

- $R_{i,t}$ and $E(R_{i,t})$ is the realized return for stock i and the model forecasted return for stock i in month t ;
- $\rho_t(R_{i,t}; E(R_{i,t}))$ is the cross-sectional correlation (the IC) between $R_{i,t}$ and $E(R_{i,t})$ in month t ;
- T is the number of months in the evaluation period;
- N_t is the number of sample stock returns being projected in month t ; and
- $\sqrt{N_t}$ is the adjustment for the breadth of the forecast in month t .

On the other hand, the Qian and Hua (2003) information ratio evaluates the model's forecasting ability with adjustments to the model's forecasting volatility.

$$\text{Qian and Hua Information Ratio} = \frac{\overline{IC}}{\sigma_{IC}} \tag{4}$$

Where:

- \overline{IC} = the average IC score of the model; and
- σ_{IC} = the volatility of the time-series ICs

EMPIRICAL FINDINGS

Variable Selection

Table 1 presents the results of the permutations of the style attributes selected under the stepwise variable selection technique for the linear models and the genetic algorithm for the nonlinear models, for each of the six overlapping training (in-sample) periods spanning the period from 01 January 1997 to 31 December 2006. Examining the identities of the style attributes included in the models periodically, it is evident that BVTP and CFTP are selected across both linear and nonlinear models in almost all periods. This observation is confirmed in Table 2 which documents the number of training periods in which each of the candidate style attributes is included by the respective linear and nonlinear models. Attributes from the value (fundamental values relative to share price) and size and return momentum categories dominate the variables selected by all models. The number of variables selected by ANN_Kalman model is substantially less than the number of variables selected by ANN_Backprop model.

Table 1: Performance Variable Selection

Table 1 documents the variables selected by the stepwise variable selection procedure (in the case of the linear models) and the genetic algorithm (in the case of the ANN models) for each of the six overlapping training periods spanning from the 01 January 1997 to 31 December 2006. Panel (A) and Panel (B) present the results of the variable selection results for the linear models, while Panel (C) and Panel (D) present the variable selection results of the ANN models.

PANEL (A) Linear_Grinfeld						
No.	1997 to 2001	1998 to 2002	1999 to 2003	2000 to 2004	2001 to 2005	2002 to 2006
1	BVTP	BVTP	BVTP	BVTP	BVTP	LSIZE
2	FOREY2	MOM12	DEBTTBVE	ROE	ROE	MOM12-1
3	CFTP	DEBTTBVE	CFTP	CFTP	CFTP	EY
4	CURRENTRATIO		ROA	DEBTTMVE	LPRICE	MOM12
5	DEBTTMVE			LPRICE	MOM12	BVTP
6	EY			*MOM12	ROA	ROE
7	LSIZE				MOM3	CFTP
8						MOM3
9						LPRICE
10						LAGLPRICE
11						

PANEL (B) Linear_QH						
No.	1997 to 2001	1998 to 2002	1999 to 2003	2000 to 2004	2001 to 2005	2002 to 2006
1	FOREY2	BVTP	BVTP	BVTP	BVTP	LPRICE
2	CFTP	MOM12-1	MOM12	MOM12-1	MOM12-1	G12MNPMargin
3	CURRENTRATIO	DEBTTMVE	DEBTTBVE	DEBTTBVE	GROWTH	CFTP
4	BVTP	EY	CFTP	**DEBTTMVE	CFTP	BVTP
5	EY	CFTP	EY		MOM1	DEBTTBVE
6	TATURNVR	FOREY2	CFTDEBT		ICBT	MOM1
7	SALESTP				G12MCPS	MOM24
8	LAGLPRICE				MOM3	MOM12
9	MOM1					MOM3
10						

Table 1: Performance Variable Selection - Continued

PANEL (C) ANN_Backprop						
No.	1997 to 2001	1998 to 2002	1999 to 2003	2000 to 2004	2001 to 2005	2002 to 2006
1	BVTP	BVTP	BVTP	BVTP	BVTP	BVTP
2	CFTCURRLIABS	CFTDEBT	CFTCURRLIABS	CFTP	CFTP	CFTCURRLIABS
3	CFTP	CFTP	CFTDEBT	DEBTTMVE	CURRENTRATIO	CFTP
4	CURRENTRATIO	DY	CFTP	DY	DEBTTBVE	EY
5	DEBTTBVE	EY	CURRENTRATIO	EG1	DY	G12MCPS
6	DEBTTMVE	G12MDPS	EG1	G24MEPS	FOREY1	G12MEPS
7	DY	G12MEPS	EY	LAGLPRICE	G12MDPS	G12MNPMargin
8	EY	G12MNPMargin	FOREY2	MOM12-1	G12MEPS	GPMargin
9	G12MDPS	GPMargin	G12MEPS	MOM24	LAGLPRICE	LSIZE
10	G12MEPS	LAGLPRICE	G24MEPS	MOM3	LPRICE	MOM1
11	G12MGPMargin	LPRICE	GROWTH	NPMargin	LSIZE	MOM12
12	LAGLPRICE	MOM1	LSIZE		MOM12	MOM6
13	LSIZE	MOM12-1	MOM12-1		MOM24	
14	MOM1	NPMargin	MOM24		MOM6	
15	MOM12-1	PAYOUT	MOM6		SALESTP	
16	MOM24	ROA	PAYOUT		TATurnover	
17	MOM6	SALESTP	ROA			
18	ROE	TATurnover	TATurnover			
19	TATurnover					
20						
PANEL (D) ANN_Kalman						
No.	1997 to 2001	1998 to 2002	1999 to 2003	2000 to 2004	2001 to 2005	2002 to 2006
1	CFTP	BVTP	BVTP	BVTP	BVTP	CFTP
2	G12MDPS	CFTP	CFTP	CFTP	CFTP	DEBTTBVE
3	PAYOUT	DEBTTBVE	DY	DEBTTBVE	EG1	LPRICE
4	BVTP	LAGLPRICE	MOM12-1	G12MCPS	G12MEPS	MOM6
5		MOM12	NPMargin	G12MEPS	G12MNPMargin	BVTP
6			TATurnover	G12MNPMargin	LAGLPRICE	
7				G24MEPS	ROA	
8				MOM24	TATurnover	
9						

Table 2: Frequency of Inclusion of Candidate Style Attributes over Examination Period

Table 2 documents the number of training periods in which each of the candidate style attributes is included each of the linear and nonlinear models.

	Linear_Grinold	Linear_QH	ANN_Backprop	ANN_Kalman
FUNDAMENTAL VALUES RELATIVE TO STOCK				
PRICE				
BVTP	6	6	6	6
CFTP	5	5	6	6
DY	0	0	4	1
EY	2	3	4	0
SALESTP	0	1	2	0
SOLVENCY AND LIQUIDITY				
CFTCURRLIABS	0	0	3	0
CFTDEBT	0	1	2	0
CURRENTRATIO	1	1	3	0
DEBTMVE	2	2	2	0
DEBTTBVE	2	3	2	3
ICBT	0	1	0	0
FUNDAMENTAL GROWTH				
G12MCPS	0	1	1	1
G12MDPS	0	0	3	1
G12MEPS	0	0	5	2
G12MGPMARGIN	0	0	1	0
G12MNPMargin	0	1	2	2
G12MSALES	0	0	0	0
G24MEPS	0	0	2	1
GROWTH	0	1	1	0
OPERATING PERFORMANCE				
GPMARGIN	0	0	2	0
NPMARGIN	0	0	2	1
PAYOUT	0	0	2	1
ROA	2	0	2	1
ROE	3	0	1	0
TATURNOVER	0	1	4	2
SIZE AND RETURN MOMENTUM				
LAGLPRICE	1	1	4	2
LPRICE	3	1	2	1
LSIZE	2	0	4	0
MOM1	0	3	3	0
MOM12	4	2	2	1
MOM12-1	1	3	4	1
MOM24	0	1	4	1
MOM3	2	2	1	0
MOM6	0	0	4	1
CONSENSUS ANALYST FORECAST				
EARNREV	0	0	0	0
EG1	0	0	2	1
FOREY1	0	0	1	0
FOREY2	1	2	1	0

Periodic and Consolidated Model Performance

The periodic performance scores of the four respective models over the 6 overlapping in-sample and corresponding out-of-sample periods are displayed in Panel (A) to Panel (F) of Table 3. All four models are evaluated by their respective in-sample Grinold (1989) information ratio and Qian and Hua (2003) information ratio. The best out-of-sample scores in each sub-period are highlighted in bold. The consolidated performance scores are presented in Table 4.

Table 3: Periodic Model Performance Results

		Performance Score	Linear_Grinold	Linear_QH	ANN_Backprop	ANN_Kalman
PANEL (A)						
1997 to 2001 2002	In-Sample	Grinold IR Score	1.7039	1.5352	1.4955	1.2330
	Out-of-Sample	Grinold IR Score	0.9170	0.8885	0.7297	1.4300
1997 to 2001 2002	In-Sample	Qian and Hua IR Score	1.2632	1.5325	0.9056	0.6215
	Out-of-Sample	Qian and Hua IR Score	0.5584	0.6333	0.5230	0.9245
PANEL (B)						
1998 to 2002 2003	In-Sample	Grinold IR Score	1.4761	1.4966	1.6121	1.4534
	Out-of-Sample	Grinold IR Score	1.8629	2.2707	1.8741	2.2972
1998 to 2002 2003	In-Sample	Qian and Hua IR Score	0.8282	0.9295	0.9679	0.8737
	Out-of-Sample	Qian and Hua IR Score	1.5470	1.7464	1.5812	1.8488
PANEL (C)						
1999 to 2003 2004	In-Sample	Grinold IR Score	1.8655	1.4158	2.1023	1.9595
	Out-of-Sample	Grinold IR Score	2.0340	2.4938	2.0893	2.4542
1999 to 2003 2004	In-Sample	Qian and Hua IR Score	1.0961	1.2480	1.3640	1.1998
	Out-of-Sample	Qian and Hua IR Score	1.1714	1.1303	1.0449	1.1365
PANEL (D)						
2000 to 2004 2005	In-Sample	Grinold IR Score	2.1144	1.8762	2.5100	2.2217
	Out-of-Sample	Grinold IR Score	1.1832	1.3822	1.3932	1.0784
2000 to 2004 2005	In-Sample	Qian and Hua IR Score	1.1583	1.2857	1.5705	1.3255
	Out-of-Sample	Qian and Hua IR Score	0.5638	1.0005	0.9121	0.5264
PANEL (E)						
2001 to 2005 2006	In-Sample	Grinold IR Score	2.1993	1.9054	2.3317	2.3289
	Out-of-Sample	Grinold IR Score	1.4999	1.6053	0.8231	0.9937
2001 to 2005 2006	In-Sample	Qian and Hua IR Score	1.3455	1.4826	1.2166	1.3116
	Out-of-Sample	Qian and Hua IR Score	1.1368	1.6951	0.3816	0.9851
PANEL (F)						
2002 to 2006 2007	In-Sample	Grinold IR Score	2.0353	1.7564	2.1686	1.6867
	Out-of-Sample	Grinold IR Score	2.2615	1.9158	0.9500	1.4660
2002 to 2006 2007	In-Sample	Qian and Hua IR Score	1.3376	1.2665	1.1679	1.0035
	Out-of-Sample	Qian and Hua IR Score	1.9808	1.6964	0.5218	1.3200

Table 4: Consolidated In-Sample and Out-of-Sample Performance Scores

This table contains the average periodic in-sample and the corresponding out-of-sample performance scores of the linear and nonlinear models developed in this paper. The best out-of-sample scores are highlighted in bold.

	Linear_Grinold	Linear_QH	ANN_Backprop	ANN_Kalman
Grinold (1989) Information Ratio (In-Sample)	1.8991	1.6643	2.0367	1.8139
Grinold (1989) Information Ratio (Out-Sample)	1.5570	1.7367	1.3426	1.6340
Qian and Hua (2003) Information Ratio (In-Sample)	1.1715	1.2908	1.1987	1.0560
Qian and Hua (2003) Information Ratio (Out-Sample)	0.9473	1.1144	0.7839	0.9762

Examining Table 3, it is observed that the nonlinear model, ANN_Kalman outperforms the linear models in the first sub-period (2002 to 2004), while the linear models outperforms the nonlinear models in the second sub-period (2005 to 2007), on a risk-adjusted basis. Also notable is the observation that the backpropagation model, ANN_Backprop has inconsistent in-sample and out-of-sample scores, indicating potential problems of model overfitting. This is also perhaps one of the major limitations of artificial neural network models in that they may typically produce superior in-sample with poor out-of-sample performance (largely due to the possibility of model overfitting). The problem of overfitting however is not evident with ANN_Kalman. According to Table 4, the linear model, Linear_QH achieves the best consolidated in-sample and out-of-sample Qian and Hua (2003) information ratio. Although the in-sample Grinold (1989) information ratio is lower for the Linear_QH, the out-of-sample Grinold (1989) information ratio is higher for this model. This is followed by the nonlinear model, ANN_Kalman, whose consolidated performance results are comparable with that of the linear model (with no large outperformance). Consistent with our prior research on global stock selection, the nonlinear model trained via the extended Kalman filter rule outperforms the nonlinear model trained via the backpropagation learning rule on the JSE over the examination period.

CONCLUSION

In this paper we evaluate 2 linear stock selection models and their nonlinear counterparts on the JSE Securities Exchange over the examination period 1997 to 2007. The linear models (which were extracted from our previous study) had the objective of maximizing the in-sample Grinold (1989) information ratio and the Qian and Hua (2003) information ratio respectively. A stepwise variable selection technique is employed to determine the optimal linear model inputs. On the other hand, the nonlinear models employ a genetic algorithm to select the best combination of input variables in the artificial neural network. The ANN models are constructed within the cascade-correlation algorithm with differing learning rules, namely the backpropagation learning rule (ANN_Backprop) and the extended Kalman filter learning rule (ANN_Kalman). A 60-for-12 dynamic moving window design is employed to update the model inputs every 12 months, based on training conducted in the prior 60 months. According to the results the in-sample and out-of-sample performance scores of the backpropagation model (ANN_Backprop) are contradictory, indicating potential overfitting problems. Consistent with our prior study on global equities, the nonlinear model trained via the extended Kalman filter rule outperforms the model trained via the traditional backpropagation learning rule. The linear model, Linear_QH is found to be the most robust model overall in the research, although the nonlinear model trained via the extended Kalman filter rule, ANN_Kalman, outperforms the linear model in the the first sub-period of the study, while the linear models outperform the nonlinear models in the second sub-period, on a risk-adjusted basis. This observation suggests further research into the complementary powers of linear and nonlinear asset pricing models.

ACKNOWLEDGEMENTS

This work is supported by the National Research Foundation (NRF) of South Africa.

AUTHOR INFORMATION

Dr. Kathleen Hodnett is currently a Research Fellow (funded by the National Research Foundation (NRF) of South Africa) and a Senior Lecturer in the School of Business and Finance at the University of the Western Cape, South Africa. She is a member of the International Institute of Forecasters (IIF) and an associate of the South African Institute of Financial Markets (SAIFM).

Dr. Heng-Hsing Hsieh, CFA is the Head of Finance in the School of Business and Finance at the University of the Western Cape, South Africa. He is a member of the International Institute of Forecasters (IIF) and a member of the South African Institute of Financial Markets (SAIFM). E-mail: ahsieh@uwc.ac.za (Corresponding author)

Prof. Paul van Rensburg is the Frank Robb Professor of Finance at the University of Cape Town, South Africa. He is the principal of Salient Quantitative Investment Management.

REFERENCES

1. Banz R W (1981), "The Relationship between Return and Market Value of Common Stocks", *Journal of Financial Economics*, vol. 9, 3-18
2. Basu S (1977), "The Investment Performance of Common Stocks in Relation to Their Price-to-Earnings Ratio: A Test of the Efficient Markets Hypothesis", *Journal of Finance*, vol. 12, no 1, 129-156
3. Cao Q, Leggio K and M Schiederjans (2005), "A Comparison between Fama and French's Model and Artificial Neural Networks in Predicting the Chinese Stock Market", *Computers and Operations Research*, vol. 32, 2499-2512
4. Cao Q, Parry M and B Leggio (2009), "The Three-Factor Model and Artificial Neural Networks: Predicting Stock Price Movement in China", *Annals of Operations Research*, Springer
5. Coats P K and L F Fant (1993), "Recognizing Financial Distress Patterns Using a Neural Network Tool", *Financial Management*, 142-155
6. Daniel K and S Titman (1997), "Evidence on the Characteristics of Cross Sectional Variation in Share Returns", *Journal of Finance*, vol. 52, no 1, 1-33
7. De Bondt W F M and R H Thaler (1985), "Does the Stock Market Overreact?", *Journal of Finance*, vol. 40, no 3, 793-805
8. De Bondt W F M and R H Thaler (1987), "Further Evidence on Investor Overreaction and Stock Market Seasonality", *Journal of Finance*, vol. 42, no 3, 557-581
9. Diamantopoulou M J (2006), "Tree-Bole Estimation on Standing Pine Trees Using Cascade-Correlation Artificial Neural Network Models", *Agricultural Engineering International: The GIGR Ejournal. Manuscript IT 06 002*, vol. 8, 1-14
10. Eakins S and R Stansell (2003), "Can Value-Based Selection Criteria Yield Superior Risk-Adjusted Returns: An Application of Neural Networks", *International Review of Financial Analysis*, vol. 12, 83-97
11. Eakins S, Stansell R and J Buck (1998), "Analyzing the Nature of Institutional Demand for Common Stocks", *Quarterly Journal of Business and Economics*, vol. 37
12. Fahlman S E and C Liebere (1990/1991), "The Cascade-Correlation Learning Algorithm", *Technical Report, CMU-CS-90-100*, Carnegie Mellon University, 1-13
13. Fama E F (1970), "Efficient Capital Markets: A Review of Theory and Empirical Work", *Journal of Finance*, vol. 25, no 2, 383-417
14. Fama E F (1991), "Efficient Capital Markets: II", *Journal of Finance*, vol. 46, no 5, 1575-1617
15. Fama E F and K R French (1992), "The Cross-Section of Expected Stock Returns", *Journal of Finance*, vol. 47, 427-465
16. Fama E F and K R French (1993), "Common Risk Factors in the Returns on Stocks and Bonds", *Journal of Financial Economics*, vol. 33, no 1, 3-56

17. Fraser E and M J Page (2000), “Value and Momentum Strategies: Evidence from the JSE”, *Investment Analysts Journal*, vol. 51, 25-35
18. Goldberg D E (1989), “*Genetic Algorithms in Search, Optimisation and Machine Learning*”, Addison-Wesley
19. Grinold R C (1989), “The Fundamental Law of Active Management”, *Journal of Portfolio Management*, vol. 15, 30-37
20. Qian E and R Hua (2003), “The Information Ratio of Active Management”, Putnam Investments
21. Haugen R A and N L Baker (1996), “Commonality in the Determinants of Expected Stock Returns”, *Journal of Financial Economics*, vol. 41, 401-439
22. Hodnett KE and H Hsieh(2012) "Application of Cascade-Correlation Neural Networks in Developing Stock Selection Models for Global Equities", *International Business and Economics Research Journal*, vol 11, no 4, 375-396
23. Hodnett K E, Hsieh H and P van Rensburg (2012a) "Payoffs to Equity Investment Styles on the JSE Securities Exchange: the Case of South African Equity Market", *International Business and Economics Research Journal*, vol 11, no 1, 19-32
24. Hodnett K E, Hsieh H and P van Rensburg (2012b) "Equity Style Payoffs and Stock Return Predictability: Evidence from the South African JSE Securities Exchange", *Journal of Applied Business Research*, vol 28, no 4, 605-618
25. Holland J H (1975), “*Adaptation in Natural and Artificial Systems*”, University of Michigan Press, Ann Arbor
26. Hung S, Liang T and V Liu (1996), “Integrating Arbitrage Pricing Theory and Artificial Neural Networks to Support Portfolio Management”, *Decision Support Systems*, vol. 18, 301-316
27. Jarvis RM and R Goodacre (2005), “Genetic Algorithm Optimisation for Pre-Processing and Variable Selection of Spectroscopic Data”, *Bioinformatics*, vol. 21, no 7, 860-868
28. Jegadeesh N and S Titman (1993), “Returns to Buying Winners and Selling Losers: Implications for Stock Market Efficiency”, *Journal of Finance*, vol. 48, no 1, 65-91
29. Jegadeesh N and S Titman (2001), “Profitability of Momentum Strategies: An Evaluation of Alternative Explanations”, *Journal of Finance*, vol. 56, no 2, 699–720
30. Kanas A and A Yannopoulos (2001), “Comparing Linear and Nonlinear Forecasts for Stock Returns”, *International Review of Economics and Finance*, vol. 10, 383-398
31. Kingdom J and K Feldman (1995), “Genetic Algorithms for Bankruptcy Prediction”, *Search Space Report, No. 09-95*, Search Space Limited, London
32. Lintner J (1965), “The Valuation of Risky Assets and the Selection of Risky Investments in Stock Portfolios and Capital Budgets”, *Review of Economics and Statistics*, vol. 47, no 1, 13-37
33. Mahfoud S and G Mani (1995), “Genetic Algorithms for Predicting Individual Stock Performance”, Proceedings of the 3rd International Conference on Artificial Intelligence Applications on Wall Street, 171-181
34. Mao K Z and S A Billings (1999), “Variable Selection in Non-Linear Systems Modeling”, *Mechanical Systems and Signal Processing*, vol. 13, no 2, 351-366
35. Markowitz H M (1952), "Portfolio Selection", *Journal of Finance*, vol. 7, no 1, 77–91
36. Motiwalla L F and M Wahab (2000), “Predictable Variation and Profitable Trading of US Equities: A Trading Simulation Using Neural Networks” *Computers and Operations Research*, vol. 27, 1111-1129
37. Rutan E (1993), “Experiments with Optimal Stock Screens”, Proceedings of the 3rd International Conference on Artificial Intelligence Applications on Wall Street, 269-273
38. Sharpe W F (1964), “Capital Asset Prices: A Theory of Market Equilibrium under Conditions of Risk”, *Journal of Finance*, vol. 19, no 3, 425-442
39. Sofge D A (2002), “Using Genetic Algorithm Based Variable Selection to Improve Neural Network Models for Real-World Systems”, Proceedings of the 2002 International Conference on Machine Learning and Applications
40. Stehle R, Bunke O and V Sommerfeld (1997), “Semiparametric Modelling of the Cross-Section of Expected Returns in the German Stock Market”, Humboldt University Berlin, Germany
41. Tobin J (1958), “Liquidity Preference as Behaviour Towards Risk”, *The Review of Economic Studies*, vol. 25, 65-86

42. Van Rensburg P and M Robertson (2003a), “Style Characteristics and the Cross-Section of JSE Returns”, *Investment Analysts Journal*, no 57, 1-10
43. Van Rensburg P and M Robertson (2003b), “Size, Price-to-Earnings and Beta on the JSE”, *Investment Analysts Journal*, vol. 58, 1-11
44. Van Rensburg and M Robertson (2004), “Explaining the Cross-Section of Returns in South Africa: Attributes of Factor Loadings: *Journal of Asset Management*, vol 4 No 5, 334-347
45. Waller B C, Marchette D J, Solka J L and E J Wegman (1996), “A Genetic Algorithm for Best Subset Selection in Linear Regression”, Proceedings of the 28th Symposium on the Interface

APPENDIX: Computation of Firm-Specific Style Attributes

No.	Descriptor	Style Attribute	Computation
(I) FUNDAMENTAL VALUES RELATIVE TO STOCK PRICE			
1.	BVTP	Book value-to-price	Book Value of Equity / Share Price
2.	CFTP	Cash flow-to-price	Cash Earnings per Share / Share Price
3.	DY	Trailing dividend yield	Ordinary Shareholders' Dividends per Share / Share Price
4.	EY	Trailing earnings yield	EPS / Share Price
5.	SALESTP	Sales-to-price	Sales per Share / Share Price
(II) SOLVENCY AND LIQUIDITY			
6.	CFTCURRLIABS	Cash flow-to-current liabilities	Net Cash Flow / Current Liabilities
7.	CFTDEBT	Cash flow-to-debt	Net Cash Flow / Total Liabilities
8.	CURRENTRATIO	Current ratio	Current Assets / Current Liabilities
9.	DEBTTMVE	Debt-to-market value of equity	Total Liabilities / Market Value of Equity
10.	DEBTTBVE	Debt-to-book value of equity	Total Liabilities / Book Value of Equity
11.	ICBT	Interest coverage before tax	Profit Before Interest and Tax / Accrued Interest
(III) FUNDAMENTAL GROWTH			
12.	G12MCPS	12-month cash holdings growth	(Current Cash Holdings per Share / Prior 12-Month Cash Holdings per Share) - 1
13.	G12MDPS	12-month dividend growth	(Current DPS / Prior 12-Month DPS) - 1
14.	G12MEPS	12-month earnings growth	(Current EPS / Prior 12-Month EPS) - 1
15.	G12MGPMARGIN	12-month gross profit margin growth	(Current Gross Profit Margin / Prior 12-Month Gross Profit Margin) - 1
16.	G12MNPMargin	12-month net profit margin growth	(Current Net Profit Margin / Prior 12-Month Net Profit Margin) - 1
17.	G12MSALES	12-month sales growth	(Current Sales / Prior 12-Month Sales) - 1
18.	G24MEPS	24-month earnings growth	(Current EPS / Prior 24-Month Earnings per Share) - 1
19.	GROWTH	Dividend growth rate	Return on Equity * (1 - Dividend Payout Ratio)
(IV) OPERATING PERFORMANCE			
20.	GPMARGIN	Gross profit margin	Gross Profit / Sales
21.	NPMARGIN	Net profit margin	Net Profit after Tax / Sales
22.	PAYOUT	Dividend payout ratio	Current Dividend per Share / Current Earnings per Share
23.	ROA	Return on assets	Net Profit Before Tax / Total Assets
24.	ROE	Return on equity	Net Profit After Tax / Ordinary Shareholders' Equity
25.	TATURNOVER	Total asset turnover	Sales / Total Assets
(V) SIZE AND RETURN MOMENTUM			
26.	LAGLPRICE	Lagged log of market price	Ln (Prior 1-Month Share Price)
27.	LPRICE	Log of market price	Ln (Current Share Price)
28.	LSIZE	Log of market capitalization	Ln (Market Capitalization)
29.	MOM1	1-month return	(Current Return Index / Prior 1-Month Return Index) - 1
30.	MOM12	12-month return	(Current Return Index / Prior 12-Month Return Index) - 1
31.	MOM12-1	Lagged 11-month return	(Prior 1-Month Return Index / Prior 12-Month Return Index) - 1
32.	MOM24	24-month return	(Current Return Index / Prior 24-Month Return Index) - 1
33.	MOM3	3-month return	(Current Return Index / Prior 3-Month Return Index) - 1
34.	MOM6	6-month return	(Current Return Index / Prior 6-Month Return Index) - 1
(VI) CONSENSUS ANALYST FORECAST			
35.	EARNREV	Earnings forecast revision	(Consensus Next EPS Forecast / Consensus Previous EPS Forecast) - 1
36.	EG1	1-year forward earnings growth	(Consensus Next EPS Forecast / Current EPS) - 1
37.	FOREY1	1-year forward earnings yield	Consensus EPS Forecast 1-year Forward
38.	FOREY2	2-year forward earnings yield	Consensus EPS Forecast 2-year Forward

Source: Table adapted from Hodnett, Hsieh and van Rensburg (2012b)

Notes

1. Linear regression is an example of direct connections from input neurons to output neurons, with no hidden neurons/layers.
2. As emphasized in Hodnett, Hsieh and van Rensburg (2012b) the data downloaded is free of look-ahead bias as DataStream International records published financial statement information as it arrives. In addition to this, stocks with less than, or equal to a turnover ratio of 0.01% are excluded for that month to ensure sufficient liquidity for sample stocks. The retention of only liquid stocks (which are not likely to be non-survivors), partially addresses the survivorship bias problem. Style attributes are winsorized and standardized according to the procedure set out in Hodnett, Hsieh and van Rensburg (2012a). Those companies involved in corporate restructuring, mergers and acquisitions or share splits are excluded as of the date of the above mentioned corporate event.
3. Genetic algorithms have been previously employed in financial applications, for example, by Mahfoud and Mani (1995) for stock selection, Rutan (1993) for portfolio selection and Kingdom and Feldman (1995) for bankruptcy prediction, to name a few.
4. Elite individuals are not involved in selection and reproduction but are automatically carried into the next generation.
5. See Hodnett and Hsieh (2012) for the mechanics of each learning rule generation.

NOTES