

A method of parameterising a feed forward multi-layered perceptron artificial neural network, with reference to South African financial markets

By ML Smith, FJC Beyers and JP de Villiers

Submission date 6 October 2014

Acceptance date 11 July 2016

ABSTRACT

No analytic procedures currently exist for determining optimal artificial neural network structures and parameters for any given application. Traditionally, when artificial neural networks have been applied to financial modelling problems, structure and parameter choices are often made a priori without sufficient consideration of the effect of such choices. A key aim of this study is to develop a general method that could be used to construct artificial neural networks by exploring the model structure and parameter space so that informed decisions could be made relating to the model design. In this study, a formal approach is followed to determine suitable structures and parameters for a Feed Forward Multi-layered Perceptron artificial neural network with a Resilient Propagation learning algorithm with a single hidden layer. This approach is demonstrated through the modelling of four South African economic variables, namely the average monthly returns on the money, bond and equity markets as well as monthly inflation. Artificial neural networks can be constructed on the aforementioned variables in isolation or, jointly, in an integrated model. The performance of a range of more traditional time series models is compared with that of the artificial neural network models. The results suggest that, on a statistical level, artificial neural networks perform as well as time series models at forecasting the returns for financial markets. Hybrid models, combining artificial neural networks with the time series models, are constructed, trained and tested for the money market and for the rate of inflation. They appear to add value to the time series models when forecasting inflation, but not for the money market.

KEYWORDS

Artificial neural network (ANN); ARIMA; financial application of ANNs; financial forecasting; money market; bond market; equity market; inflation

CONTACT DETAILS

Dr FJC Beyers, Department of Insurance and Actuarial Science, University of Pretoria, Private Bag X20, Pretoria; Tel: +27(0)12 420 4419; Email: conrad.beyers@up.ac.za

1. INTRODUCTION

1.1 The aim of this paper is to present a methodology for using Artificial Neural Networks (ANNs) to construct a forecasting system for financial data. The method also includes the estimation of parameters required for ANNs to function efficiently. The parameter estimates include the number of input and hidden neurons, the number of training epochs, and the learning rate increase and decrease. The financial data considered are the returns on the money, bond and equity markets and the rate of inflation for the years 1975 to 2010. Furthermore, in the instance of inflation, additional data for the period from 2011 to 2013 have been taken into consideration. The forecasting system is used to forecast the returns on the markets and the rate of inflation for periods of one and three months in advance. An additional period of twelve months is considered for the money market. Using ANNs, a forecasting model is constructed for each market in isolation. These isolated systems are then combined into an integrated system that allows inter-market relationships to influence the forecasts. The forecasting systems are then compared with a range of time series models. Finally, hybrid models are constructed and their performance evaluated.

1.2 ANN modelling is a dynamic methodology that is part of the class of soft computing approaches that form a considerable portion of modern artificial intelligence techniques (Russell & Norvig, 2010). The name 'Artificial Neural Network' is derived from the origin of the model, namely, that the model attempts to replicate the functioning of the human brain (Zhang, Patuwo & Hu, 1998). However, the brain is an extremely complex organ and virtually impossible to replicate. An ANN is therefore limited to replicating the functioning of the human brain for a very specific task. Furthermore, ANNs are rational agents, which means that they 'make the best decision' based on the information available to them (Russell & Norvig, 2010).

1.3 Although ANNs appear different from statistical methods, the methods are very closely related. ANNs identify and use non-linear relationships within data, and are therefore appropriate for real-world problems, as these are also non-linear in nature (Zhang, Patuwo & Hu, 1998). However, ANNs have often been associated with 'black box' models because the full interpretation of the internal functioning of ANNs is currently not possible. The results obtained in this study from ANNs support the proposition that, in some instances, they offer better accuracy than traditional statistical models (Huang, Zhu & Siew, 2006). Furthermore, the technique is immature, leaving room for discovery and improvement. Another aspect unique to ANNs is their ability to learn and evolve with experience. As these models can be applied in dynamic environments without the need of regular redesign (Tan, undated),

they are prime candidates for use in the new environments. Finally, ANNs perform well with vaguely defined problems, unfinished data, inaccuracies and uncertainties (Korol, 2013).

1.4 There are various types of ANNs and associated learning algorithms. This paper focuses on Feed Forward Multi-layer Perceptron (FF MLP) (Huang, 2009; Wilamowski, 2009), with implementation of the Resilient Propagation (RPROP) (Allard & Faubert, 2004) learning algorithm.

1.5 ANNs have been applied in a number of areas in the financial markets, including forecasting forward interest rates (Bouqata, Bensaid & Palliam, 1999), estimating general insurance reserves (Braun & Lai, 2005), evaluating credit risk (Fei & Zhigang, 2008), detecting credit card fraud (Sahin & Duman, 2011), analysing stock exchange movements (Abhishek et al., 2012; O'Connor & Madden, 2006), forecasting commodity prices (Kohzadi et al., 1996) and foreign exchange rates (Tan, undated), and predicting credit union distress and bankruptcy (Korol, 2013; Tsai & Wu, 2008; Saha, 2009; Lacher et al., 1995; Tan, undated).

1.6 For implementation considerations, several qualities of ANNs are advantageous in comparison with more traditional time series models. These qualities include easy implementation and integration into existing models, efficient handling of random variation within the datasets, versatility and hence applicability to different types of problems, easy updating, the handling of a large number of inputs without leading to complex parameter estimations, robustness in terms of forecasting accuracy, and applicability to any computing platform.

1.7 ANNs do have some disadvantages. Mainly, these have to do with the lack of developed theoretical optimisation techniques, the possibility of overfitting, lack of understanding of their internal functioning, the difficulty of understanding the relationship between the inputs and outputs, and the need for significant computational power for large datasets.

1.8 A natural extension of ANNs is the topic of deep learning. Deep learning, in its simplest form, consists of ANNs with many hidden layers, allowing the models to discover intricate structure in large datasets. ANNs were largely forsaken by the machine-learning community in the late 1990s, because of the commonly-held belief that feature extraction with no prior knowledge was infeasible. Further, training algorithms such as error minimisation, often failed to identify global minima (by getting stuck in local minima), leading to unreliable fitting to data (LeCun, Bengio & Hinton, 2015).

1.9 Since 2006, deep learning has been increasing in popularity. Several factors have contributed to this revival:

- In practice, local minima do not have a detrimental effect on performance for large networks;
- Unsupervised methods have been used to pre-train the networks, improving the feasibility of feature detection; and

— Graphical processing units have been used to train the networks, making training of large networks possible (LeCun, Bengio & Hinton, 2015).

1.10 This article does not focus on deep learning, but on the building blocks used to develop a specific type of deep learning model. For more information on deep learning, see LeCun, Bengio & Hinton (2015).

2. THEORETICAL BACKGROUND

2.1 Origin of Artificial Neural Networks

2.1.1 According to Charles Darwin, the species most adaptable to change is the species most likely to survive. A trait is shared by the animals able to adapt: the ability to learn. It is likely that the earth would be barren and devoid of life if animals had not this ability (Schmidt, 2000).

2.1.2 The brain is the organ from which the ability to learn emanates. Learning is performed through a highly specialised network of cells (neurons) located in the brain. Neurons are considered to be the building blocks of the central nervous system in animals, comparable to processors in computers, and perform similar tasks.

2.2 Artificial Neuron

2.2.1 On a small scale, ANNs replicate the rational functioning of the human brain. Figure 1 is a diagram of a single artificial neuron, and is referred to during the explanation of ANNs.

2.2.2 Inputs allow data into the artificial neuron. These data could have been carried over from the original dataset, or are outputs of the preceding neurons. Each input variable (x_{ij}) represents a variable from the dataset, or a single output from a preceding neuron.

2.2.3 Every layer is fully connected to both its preceding and its succeeding layer (for more information on layers refer to section 2.3). Each connection has a level of strength, which is the ‘weight’ of the connection. Weights are conventionally denoted by (w_{ij}), where ‘ i ’ is the neuron in the preceding layer and ‘ j ’ is the neuron in the current layer. ANNs learn and adapt in dynamic environments through the manipulation of these weights.

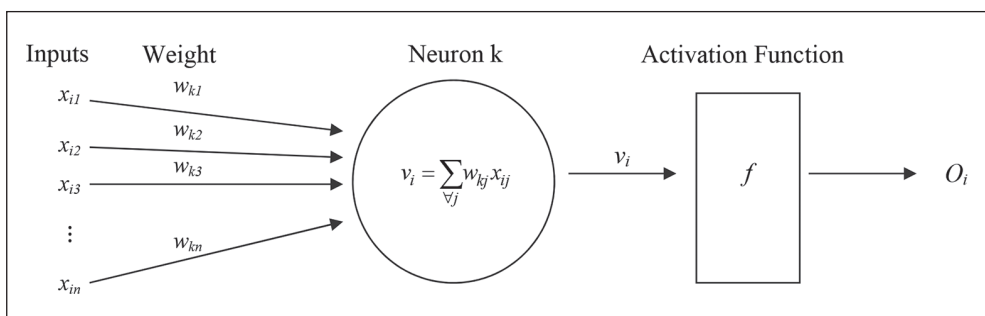


FIGURE 1. An artificial neuron

2.2.4 The energy of a neuron is determined in the neuron itself. The energy is defined as the sum of all the inputs multiplied by their associated weights and is denoted as v_i in Figure 1.

2.2.5 The energy generated in the neuron forms an output using an activation function. Any function can be used as an activation function, although it is common to use a sigmoid function. The logistic function or the hyperbolic tangent are common choices (Lacher et al., 1995). In this study, we used the hyperbolic tangent function $f: \mathbb{R} \rightarrow (-1,1)$, defined by

$$f(x) = \frac{2}{1 + e^{-x}} - 1 \quad (1)$$

2.2.6 The output represents the results of the system, which consist of either a single value or a vector of values. In the case of regression problems that require the scaling of inputs, the scaling process must be reversed to obtain relevant output. This should be done before the output is finalised. The output is denoted by O_i in Figure 1.

2.3 Artificial Neural Network Structure

2.3.1 A MLP is the oldest structure of ANNs and consists of an input layer, possibly several hidden layers, and an output layer. Each layer consists of artificial neurons and interconnects with the preceding and succeeding layer. No connections bypass a layer, that is, there are no direct connections from the input to the output layer. Furthermore, the number of hidden layers and the neurons within those layers are specified in advance. There is no theoretical method of determining the correct number of layers and neurons within those layers. Figure 2 provides a diagram of the network with two hidden layers. For further details refer to Wilamowski (2009).

2.3.2 A Feed Forward network is the most common type of ANN. This network allows a unidirectional flow of information, i.e., from input to output. Information is not allowed to flow backwards through the network, i.e., from output to input. Figure 3 depicts the information flow through the connections of a FF MLP (Huang, 2009).

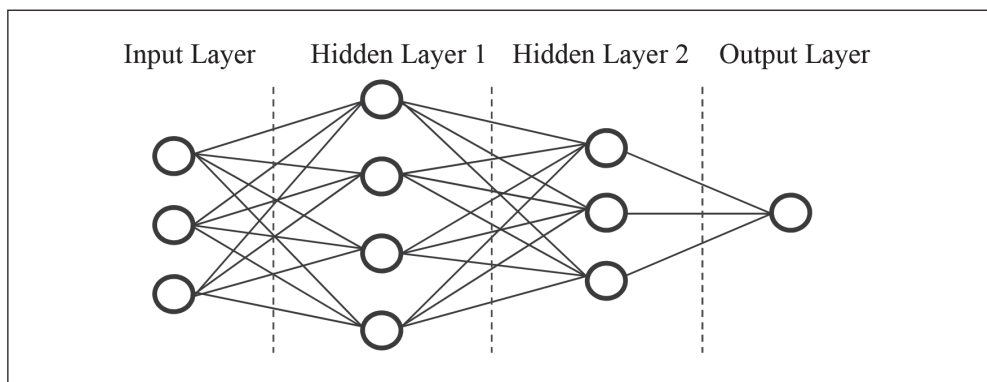


FIGURE 2. MLP (3 input neurons—4 hidden neurons—3 hidden neurons—1 output neuron)

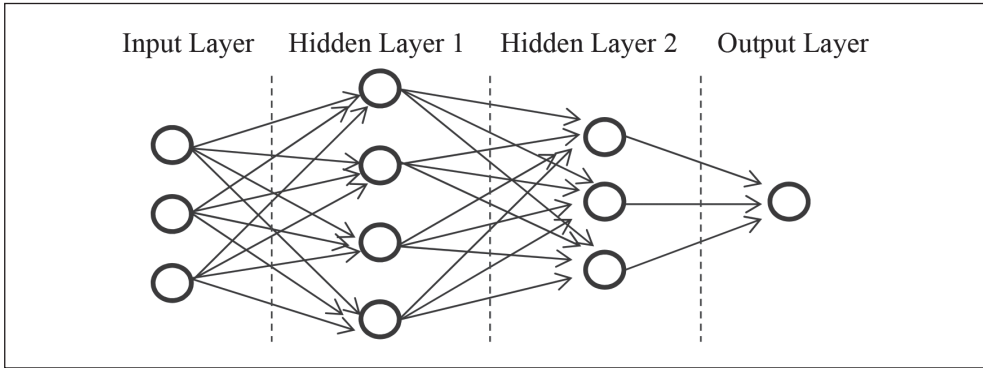


FIGURE 3. Feed Forward MLP (3—4—3—1)

2.4 Learning Algorithm

2.4.1 The learning algorithm implemented for this study is referred to as Resilient Propagation (RPROP). The RPROP algorithm was chosen because it improves on other learning algorithms such as Back Propagation (BP), Quick Propagation and Super Self-Adapting Back Propagation (Super SAB) (Riedmiller & Braun, 1993). One strong characteristic of RPROP is that only the direction of the gradient of the error measure (sign of the partial derivative) is used to manipulate the weights. Additionally, RPROP does not use the magnitude of the error to update the weights as BP does; consequently, all the areas of the system are trained evenly and more efficiently. This results in RPROP being superior to most first-order training algorithms with regard to training time and accuracy. Each connection is assigned a unique learning rate, which localises learning and allows weights to be updated in accordance with their contribution to the system error. Consequently, the vagaries caused by a single generalised learning rate are removed and the learning rate is directly and solely responsible for the size of the weight update.

2.4.2 Each weight is assigned a random learning rate (LR) when the system is initialised and these rates increase or decrease exponentially in order to minimise the error of the system. The gradient of the error function at the previous and current updates determines whether the learning rates increase or decrease. The gradient of the error function is determined, in turn, by its partial derivative with respect to a specific weight. If there is a change in the sign of the partial derivative of the error measure (gradient direction of the error surface), with respect to the weight being considered, the optimal minimum has been missed. The result of missing the optimal minimum is a decrease in the learning rate. If the sign of the partial derivative remains the same, the learning rate is increased. This technique requires that there be bounds on the learning rates. By convention, these limits have been 1×10^{-6} and 50 respectively (Allard & Faubert, 2004). RPROP requires two parameters to be estimated, an exponential learning rate increase and a decrease. For more information on RPROP, refer to Allard & Faubert (2004) and Riedmiller & Braun (1993).

2.4.3 An increase in the learning rate reduces the time until the convergence of the ANN. This increase is achieved by increasing the size of the learning step if the sign

of the gradient remains constant over two consecutive weight updates. Two consecutive identical signs indicate that the system is approaching a minimum error. By increasing the step size, the minimum error will be approached with an increased speed, as is illustrated by Figure 4. The optimal learning-rate increase parameter is difficult to determine and there is little theoretical study available on it. There is a danger that the learning rate increase could be significantly larger or smaller than the optimal parameter. An excessively small learning rate increase requires a large number of training epochs for the ANN to converge. A large learning rate increase causes the system to ‘overshoot’ the minimum error, resulting in a significant number of oscillations and increased training time.

2.4.4 The decrease in the learning rate is responsible for the reduction in the learning step size if the system ‘overshoots’ the minimum error, as is shown in Figure 5. If the learning rate decrease is not present, the system will begin to oscillate and the error will become trapped at a certain point (as depicted by the dashed blue line in Figure 5). As with the

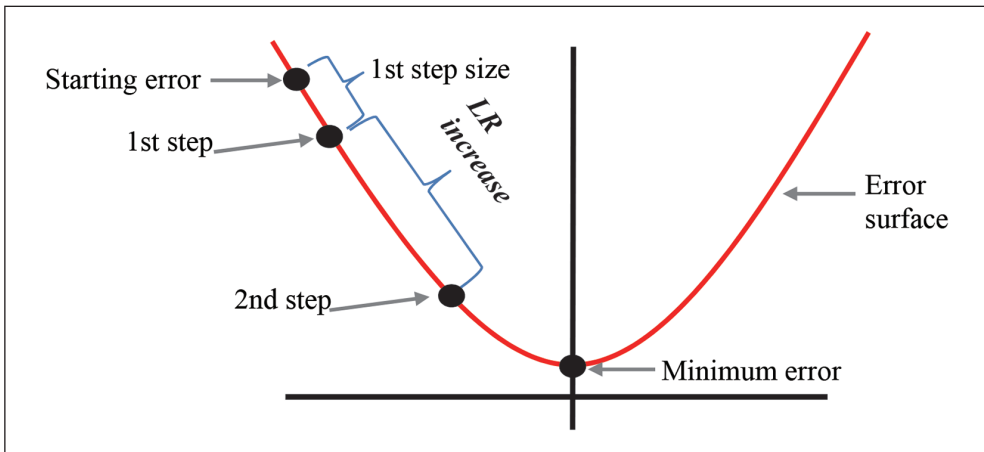


FIGURE 4. Learning rate increase

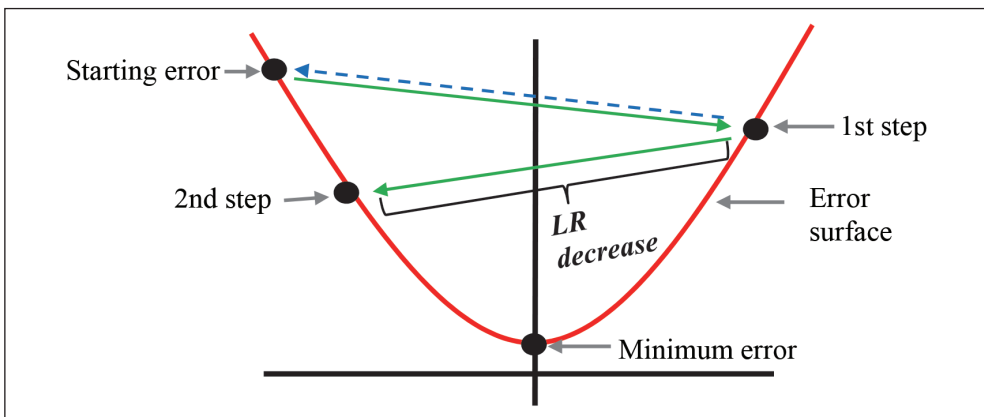


FIGURE 5. Learning rate decrease

learning rate increase, the learning-rate decrease parameter is difficult to estimate accurately. An excessively small learning rate decrease results in a significant number of oscillations before the minimum is obtained. A large learning rate decrease requires a large number of training epochs and a large amount of computing power to achieve the convergence of the system.

3. METHOD OF ARTIFICIAL NEURAL NETWORK CONSTRUCTION

3.1 When constructing an ANN of a given type, modelling decisions need to be made in relation to the structure and parameters of the network. This includes the number of input, hidden and output layers and the associated number of neurons in each layer. Further design choices include the neuron activation function and the initial weights, as well as the learning algorithm and associated parameters. The learning algorithm, which can be chosen from several candidate algorithms, is itself parameterised. The FF MLP using the RPROP learning algorithm requires several parameters to be estimated, as described in Table 1.

TABLE 1. Parameters required for an ANN

Parameter	Description
Number of initialised structures/ number of times the structure is initialised	The weights and learning rates of the system are initialised randomly. This is the number of times the ANN is initialised and subsequently trained.
Learning increase/decrease	The speed of learning. High rates increase speed; however, they may lead to larger errors. Low rates increase accuracy but also increase computing time.
Training epochs (epochs)	The number of times the system is trained. High values reduce the error of the ANN over the training set of data; however, this may lead to overfitting in some instances.
Input neurons/lagged observations (lags)	The number of lagged observations used in the forecast process.
Hidden neurons	The number of neurons, which are found in the hidden layers, used for processing the information.

3.2 Parameter Estimation Method—Artificial Neural Networks

3.2.1 The required parameters are estimated through a simulation approach. First, the number of epochs is estimated, which ensures that every ANN structure considered will obtain its minimum error over the testing dataset during training (refer to section 4.3 for the training and testing set specifics). Several combinations of parameters for the ANNs are considered in turn. Furthermore, the ANNs are trained over several randomly generated initial weight value vectors to determine the average ANN performance over several Monte Carlo simulation runs (as specified in ¶3.7). Each ANN is initialised and trained several times and the average mean squared error (MSE) is determined, which limits the effect of random fluctuations. This leads to a multivariate distribution of the MSE.

3.2.2 Using the established root mean squared error (RMSE) distribution, the learning rate increases and decreases are estimated. This is achieved by constructing a three-dimensional surface, with the axes consisting of MSE, the learning rate increase, and the learning rate decrease respectively. The average MSE over the different combinations of input and hidden neurons is determined and used to construct the surface. It is important to note that the average MSE does not explicitly allow for outliers in the data and the result could therefore be skewed. The scope of this study did not include extensive testing to assess the effect of outliers on the results.

3.2.3 The number of neurons in each layer is chosen by searching over a discrete space of possible input and hidden neurons. The search criterion is the minimisation of the MSE over the testing dataset. Only the testing set is used as the choice of number of neurons could be distorted through overfitting to the training set. The MSE surface over the training and testing set is plotted for the chosen learning rate increase and decrease parameters. The learning rate parameters are determined through a search over a discrete space of learning rate increases and decreases. Here the search criterion is the minimisation of the MSE over both the training and testing datasets. These surfaces are analysed and a combination of input and hidden neurons is chosen by selecting parameters that minimise the trade-off between errors over both surfaces. This decision has subjective components. In this instance it is important not to base the decision purely on the results obtained from the training set, as this would lead to overfitting. Similarly, basing the decision purely on the results over the testing set would cause the ANN to fit only the testing set. Both the scenarios described above would reduce the generalisation ability of the ANN, leading to poor forecasting accuracy. Examples of such error surfaces over the training and testing datasets respectively are provided in Figures 6 and 7.

3.2.4 Once all the parameters are estimated, the ANN is trained for an increased number of epochs, recording MSE values that correspond to different numbers of training

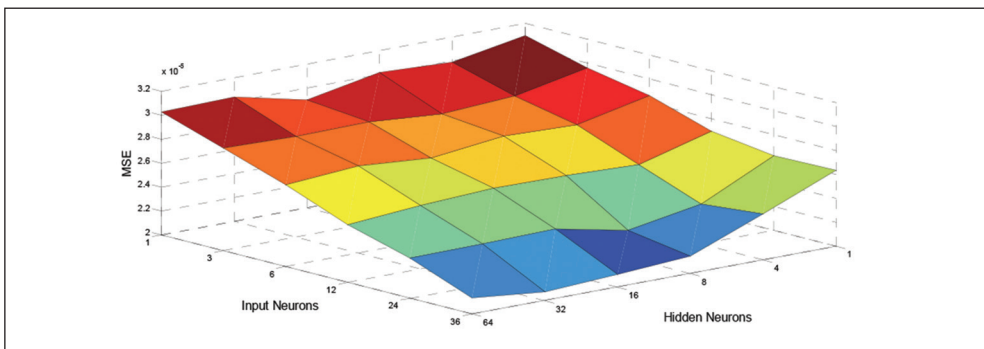


FIGURE 6. MSE surface over training dataset for inflation one-step isolated ANN application¹

¹ The colours of the surface provide an indication of the absolute level of the error surface. Blue represents a point of low error, where red indicates a high error point. This scale is used consistently throughout this paper.

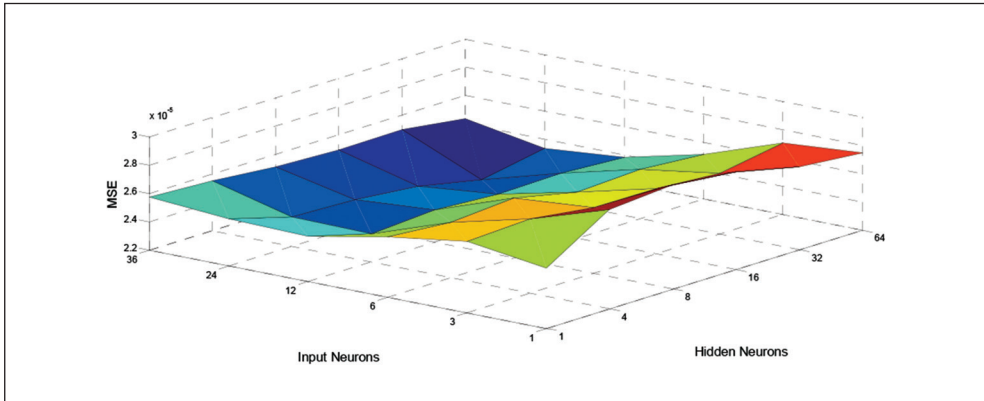


FIGURE 7. MSE surface over testing dataset for inflation one-step isolated ANN application

epochs. The number of epochs that correspond to the lowest MSE is then chosen. By using this number of training epochs, the likelihood of overfitting within the model should be reduced.

3.2.5 Limiting values are determined for each parameter mentioned in ¶3.1, simplifying the estimation process. The estimations are explained in ¶¶3.2.6–3.2.10.

3.2.6 Autocorrelation plots are used to determine the maximum number of past observations that has significant correlation to the current observable value. This provides a maximum for the number of inputs. A limiting case for the number of hidden neurons needs estimation and a number is chosen in order to be sufficient to capture the necessary relationships.

3.2.7 Limiting values for the learning rate increase and decrease parameters are required for the estimation processes. These need to be larger than zero yet smaller than one.

3.2.8 Two additional variables are used to explain seasonality in the datasets. The only dataset containing seasonality is inflation, which includes characteristic annual seasonality. The variables used to capture the seasonality are determined using the following process. Each month is assigned a number ranging from 1 to 12 (1 = January, 2 = February, etc.). Each number is divided by 12, which provides the proportion of 12 each period represents. This number is presented to a cosine and sine function to produce two variables; the variables are used to capture any seasonal trends in the data and are added to the dataset as additional explanatory variables.

3.2.9 The MSE is used as the error measure because the learning algorithm is derived from the MSE.

3.2.10 The data are normalised by dividing each data point by the largest absolute observation. This method ensures that the observations are scaled into the interval $[-1; 1]$.

3.2.11 Matlab R2011a software was used to code the ANNs used in this experiment and Microsoft Excel 2013 was used for the analysis of several results.

3.3 Time Series Model Construction

3.3.1 Autoregressive integrated moving average (ARIMA) models are used to compare the forecasting ability of the ANNs and are constructed in the statistics program SAS.

3.3.2 Several best-fit ARIMA models are constructed for the same dataset as that used for the ANNs. The chosen ARIMA model is the model that achieves the lowest MSE over the testing set of data. The time-series forecasting system in SAS only allowed future n-step forecasts over the testing set of data from a static point in time.

3.4 Hybrid Model Construction

3.4.1 Models that combine the time series models and ANNs are referred to as hybrid models. In this study, such models are constructed for the inflation rate and the money market, over forecast periods of one month and three months, to be comparable with the other models.

3.4.2 The time series models are applied to the raw data to detect possible trends. The residuals produced are presented to the ANNs, which stripped the remaining relationships from the data. Therefore, the forecasts from the time series models and the forecasts of the residuals from the ANNs are combined and they represent the forecasts generated by the hybrid model.

3.5 Method of Comparison

3.5.1 The ANNs, time series models and hybrid models are compared by constructing comparison intervals as described below. Noting the risk of bias in results attributable to random variation, the RMSE of both of the forecasts generated by each model and of the resulting residual series are considered. This is carried out over the testing set, instead of the training set, as overfitting, which may be present in the training set, could skew the results.

3.5.2 Comparison intervals are constructed by increasing or decreasing the RMSE of each model. The magnitude of these increases and decreases are chosen to be a single standard deviation of the residuals of the forecasts that are generated by the model over the testing set (which contained 100 data points). The form of the interval is given by

$$[\max(RMSE - \sigma_{residuals}, 0); RMSE + \sigma_{residuals}] \quad (2)$$

Let ε_t be the residual of forecast t (1), i.e.

$$\varepsilon_t = (\hat{X}_t - X_t) \quad (3)$$

where \hat{X}_t represents the forecast at time t and X_t the actual value at time t . A residual series ε is created by the forecasts, and is given by

$$\varepsilon = (\varepsilon_1, \varepsilon_2, \dots, \varepsilon_{100}). \quad (4)$$

The standard deviation $\sigma_{residuals}$ of this series is given by

$$\sigma_{residuals} = \sqrt{\frac{1}{99} \sum_{t=1}^{100} (\varepsilon_t - \bar{\varepsilon})^2} \quad (5)$$

where

$$\bar{\varepsilon} = \frac{1}{100} \sum_{t=1}^{100} \varepsilon_t. \quad (6)$$

The lower limit of this interval has a minimum of zero because the RMSE is non-negative by definition. These intervals are determined for all the models considered. No overlap between their comparison intervals would indicate that the performance of one model is significantly different compared with that of the other. This method of comparison takes into account the RMSE of the forecasts and the volatility within the residuals of the forecasts.

3.5.3 The method of comparison explained in the preceding paragraph is used because the underlying distributions of the forecasts, RMSE, and forecast residuals are not known. Furthermore, the standard deviation of the residuals is used to increase or decrease the RMSE of the forecasts, instead of the mean error of the residuals, because the mean error of the residuals is close to zero in all the instances. This result is a by-product of the model-fitting process. It would be necessary to use additional tests in instances where the test described above indicates a significant difference in the performance of the models. However, if no such instance is found, it is unlikely that additional tests would produce different conclusions.

3.6 Cross-Validation of Models

3.6.1 The general forecasting accuracy of the models (isolated and integrated ANNs and the time series models) is determined by cross-validation. In addition, the results are used to compare the ANNs with the time series models and to illustrate the general methodology. The performance of the two modelling approaches (ANNs and time series) are evaluated over sets of data independent of the training data in order to allow appropriate comparison to be made. Evaluation over independent sets of the data reduces the effect that random variation might have on the comparison of the two modelling approaches.

3.6.2 Cross-validation of the models is only performed on the one-month inflation forecasts and is completed for both the isolated and the integrated ANN. In further investigations, the scope could be expanded to include the money market and hybrid models, and a forecast period of three months.

3.6.3 The data is divided into five sets, each with 100 observations, and each set is further divided into a training set and a testing set. The training set consisted of the first 80 observations, while the testing set consisted of the remaining twenty.

3.6.4 To establish the performance of the models, a 'confidence interval' of each error is constructed and compared. These intervals are given by

$$[\overline{RMSE}_i - \sigma_{RMSE_i}; \overline{RMSE}_i + \sigma_{RMSE_i}] \quad (7)$$

for each model i . In this study a model is regarded as statistically superior to another model if the upper limit of its confidence interval (7) is smaller than the lower limit of the other model's interval. The confidence intervals are constructed using the RMSE of each model over five different sub-divisions of the inflation dataset which allows comparison between each modelling approach. The mean and standard deviation of the RMSE are calculated. The interval construction process is described below (8,9):

$$\overline{RMSE}_i = \frac{1}{5} \sum_{t=1}^5 RMSE_i(t) \quad (8)$$

where $RMSE_i(t)$ is the RMSE of model i for dataset t . Further,

$$\sigma_{RMSE_i} = \sqrt{\frac{1}{4} \sum_{t=1}^5 (RMSE_i(t) - \overline{RMSE}_i)^2} \quad (9)$$

where σ_{RMSE_i} is the standard deviation of the RMSE for model i .

3.7 ANN Initialisation

The weights and the learning rate of the ANNs used in each application must be initialised before the ANNs can be trained. The weights and learning rate are initialised using a normal distribution with a mean of zero and a standard deviation of 0.001. The modelling outcomes of the experiment are not sensitive to the initialisation values, but they impact the training time required.

4. DATA FOR CASE STUDY

4.1 Data Characteristics

4.1.1 The financial data used in this experiment was collected from the capital markets of South Africa and relates to the money, bond and equity markets, as well as inflation. The data was obtained from two capital market history studies (Firer & McLeod, 1999; Firer & Staunton, 2002), which have identical inflation, money market and bond market returns but differing average equity returns² that date back to 1925. The data available on the bond and money markets date back to 1929 and 1945 respectively, and on the rate of inflation to 1939. All returns recorded are inclusive of re-invested dividends or interest. The data used are average monthly readings of each market and inflation from 1975 to 2010, to maintain consistency of approach across all the markets considered.

2 The difference in the average equity returns is due to differing focus of the two studies. Study 1 (Firer & McLeod, 1999) focuses on returns driven by commercial and industrial equities. Study 2 (Firer & Staunton, 2002) focuses on the inclusion of mining equities. From 1960, the returns on equity are common.

4.1.2 The Consumer Price Index (CPI) of South Africa is used to measure inflation. Readings are produced by the Central Statistical Service. CPI is widely used as a measure of inflation since it relates to the change in price over time of a basket of consumer goods. This basket is constructed using a sample of 30000 household buying patterns and represents the average basket of goods bought by a household. The dataset consists of CPI readings on a monthly basis over a range of 36 years (or 432 months) from 1975 to 2010.

4.1.3 There were and still are no official indices tracking the overall performance of the money market. In this case, an index has to be chosen which represents the average return from the money market. There are several possible instruments that can be used, ranging from bankers' acceptances and treasury bills to negotiable certificates of deposit (NCD). NCDs were chosen as these historical rates are not distorted by the various investment requirements implemented in different industries from 1960 to 2010. Further, during the 1990s, 40% to 50% of the money market's capitalisation consisted of NCDs. Finally, NCDs were issued for varying durations, up to one year, providing greater information on the yield curve of the money market. Other instruments were only issued for periods of three months. The observations used in this investigation range over 36 years (432 months) from 1975 to 2010.

4.1.4 The period considered (1975–2010) includes the construction of the JSE-Actuaries All Bond Index in 1986, used to describe the performance of the bond market from then on. For the period 1975–1986 the performance on long-term bonds is used as a proxy of the bond market's performance. Inconsistencies in the data are to be expected from 1986 onwards. The All Bond Index has two constituents, a capital section and an income section. These sections are widely known as the Price and the Yield Index. The Price Index is the market capitalisation of a portfolio of bonds. The volume of each bond is weighted in proportion to the amount held privately. The Yield Index is constructed by considering the closing yield on the JSE Gilt floor, which provides a daily index. The constituents comprising the portfolio of bonds are reviewed quarterly. The income accrued on the bonds is assumed to be smooth and the actual discrete income payments are ignored. This ensures a smooth progression of indices over a single year. The monthly yield and price are calculated on the last day of each month. This implies that they are date-dependent and may be sensitive to random fluctuations.

4.1.5 The data used to determine returns from the equity market are based on the available JSE-Actuaries All Share Index, from 1978 to 2010. Prior to this, the *Rand Daily Mail* Industrial Index (RDM100) is used. The RDM100 constituents were changed infrequently by the editors of the *Rand Daily Mail*. Further, limited information is available on the methodology behind the calculation of the RDM100. The methodology behind the calculation of these constituents improved in 1995 and now covers the full range of equity investments. The index consists of a price and income index.³

3 The price index is based on a market capitalisation weighted average of its components where the weights are given by market capitalisation. Readings were taken every two minutes while trading was possible. The income portion of the index is the dividend yield on the constituents. It is assumed that dividends are received evenly throughout the year, which does not hold true. However, this is a reasonable assumption based on a study conducted which investigated the timing of dividend

4.1.6 The dataset is analysed and the characteristics noted, as indicated in Table 2. From the investigation the risk-return profile of returns are broadly as expected from the theory.

TABLE 2. Characteristics of financial data (monthly returns)

	Inflation	Money market	Bond market	Equity market
Minimum	-1.15 %	0.37%	-14.46%	-29.71%
Maximum	4.32 %	1.94%	11.41%	17.85%
Mean	0.80 %	1.01%	1.07%	1.66%
Standard deviation	0.60 %	0.35%	2.44%	6.25%
Range	5.46 %	1.56%	25.86%	47.55%

4.2 Correlation between Datasets

4.2.1 The correlation between the markets is determined and analysed. The calculated correlations are provided in Table 3.

TABLE 3. Pearson's Rho and Spearman's Rho correlation matrix

	Pearson's Rho correlation matrix				Spearman's Rho correlation matrix			
	Inflation	Money	Bond	Equity	Inflation	Money	Bond	Equity
Inflation	1.00	0.20	-0.07	0.02	1.00	0.24	-0.13	0.02
Money	0.20	1.00	0.14	-0.05	0.24	1.00	0.15	-0.07
Bond	-0.07	0.14	1.00	0.30	-0.13	0.15	1.00	0.21
Equity	0.02	-0.05	0.30	1.00	0.02	-0.07	0.21	1.00

4.2.2 The correlation matrices indicate that the returns on the money market are related to the rate of inflation. The correlation is approximately 0.2. A slightly negative correlation exists between inflation and the return on the bond market, but, as this is close to zero, it is not expected to have a significant effect on the individual markets. It is noted that the rank correlation suggested a greater relationship between the two variables than the linear correlation. The rate of inflation and the return on the equity market are unrelated, as is expected, as the equity market is driven by supply and demand, and factors other than inflation. The correlation measures indicated little correlation between either the returns on the money and bond markets, or the returns on the money and equity markets. However, a significant relationship, and the greatest linear correlation measure throughout all the datasets

payments during the year (Firer & McLeod, 1999). The data are converted to monthly returns based on the price index. The index is assumed to be purchased at the beginning of the month and sold at the end. The month's dividend is assumed to be received half-way through the month and is reinvested in the index at its average price over that specific month. The return on the index is determined by dividing the end-of-month value (including reinvestment income) by the corresponding value at the beginning of the month.

(0.3), exists between the returns on the bond and equity markets. The correlation is relatively, but not absolutely, high, indicating that important external explanatory variables are not included in the data.

4.3 Training and Testing Datasets

4.3.1 The training set used data ranging from January 1975 to August 2002, and contained 332 observations, representing the majority of the observations. The training set is used to fit the time series models and to train the ANNs.

4.3.2 The testing set used data ranging from September 2002 to December 2010, and consisted of 100 observations not contained within the training dataset.

5. APPLICATION SPECIFIC ARTIFICIAL NEURAL NETWORK DESIGNS

5.1 The designs of the constructed ANNs are presented in Figures 8, 9 and 10. Figure 8 represents the design of the ANN that is used to forecast the rate of inflation in isolation. Figure 9 is a diagram of the ANN that is used to forecast the return on each market, while Figure 10 is a diagram of the integrated ANN.

5.2 For the limiting cases of the parameter estimates (as discussed in section 3.2) the partial autocorrelation plots suggested that observations pre-dating a three-year period would have a negligible influence on the forecast. A maximum of 36 input neurons are therefore considered, as this number corresponds with a three-year period. The number of input neurons considered and tested is increased exponentially in order to limit computing time. The exponential increase is motivated by the fact that the addition of a single neuron has a decreasing significance as the total number of neurons increases. A maximum of 64 hidden neurons is sufficient to capture all the relationships in the data. The number of hidden neurons is varied in an exponential manner and, as in the instance of the input neurons, the same

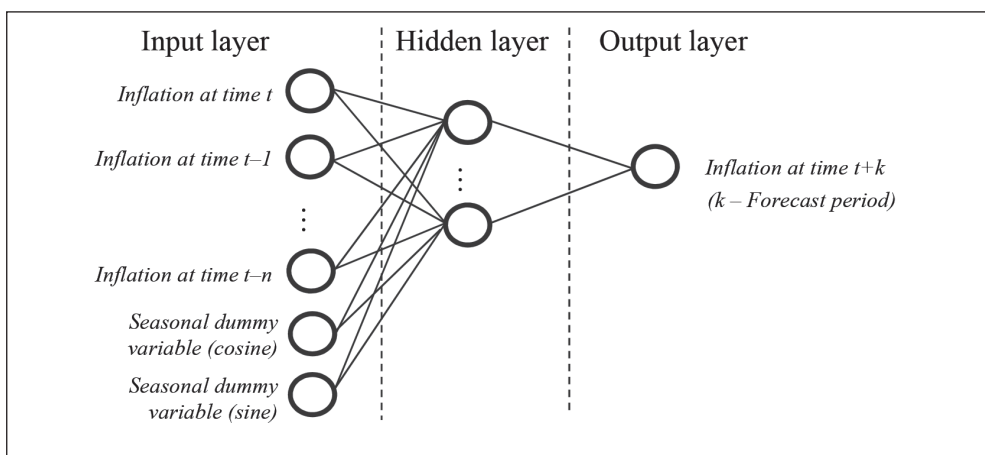


FIGURE 8. Design of inflation forecasting isolated ANN

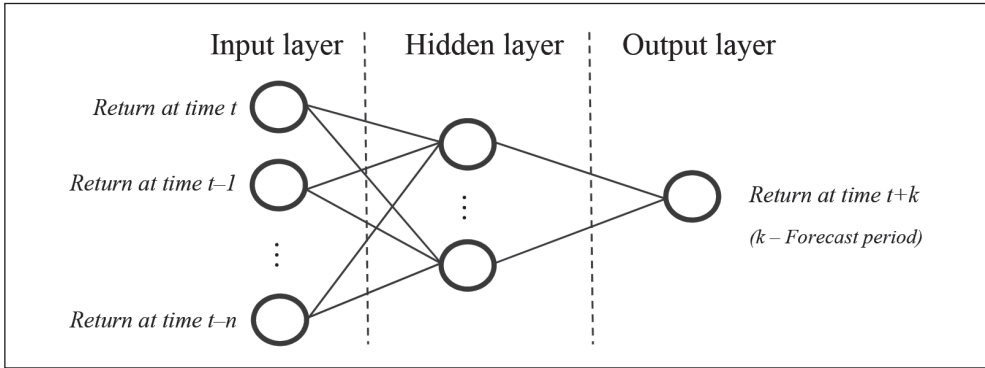


FIGURE 9. Design of return forecasting isolated ANN

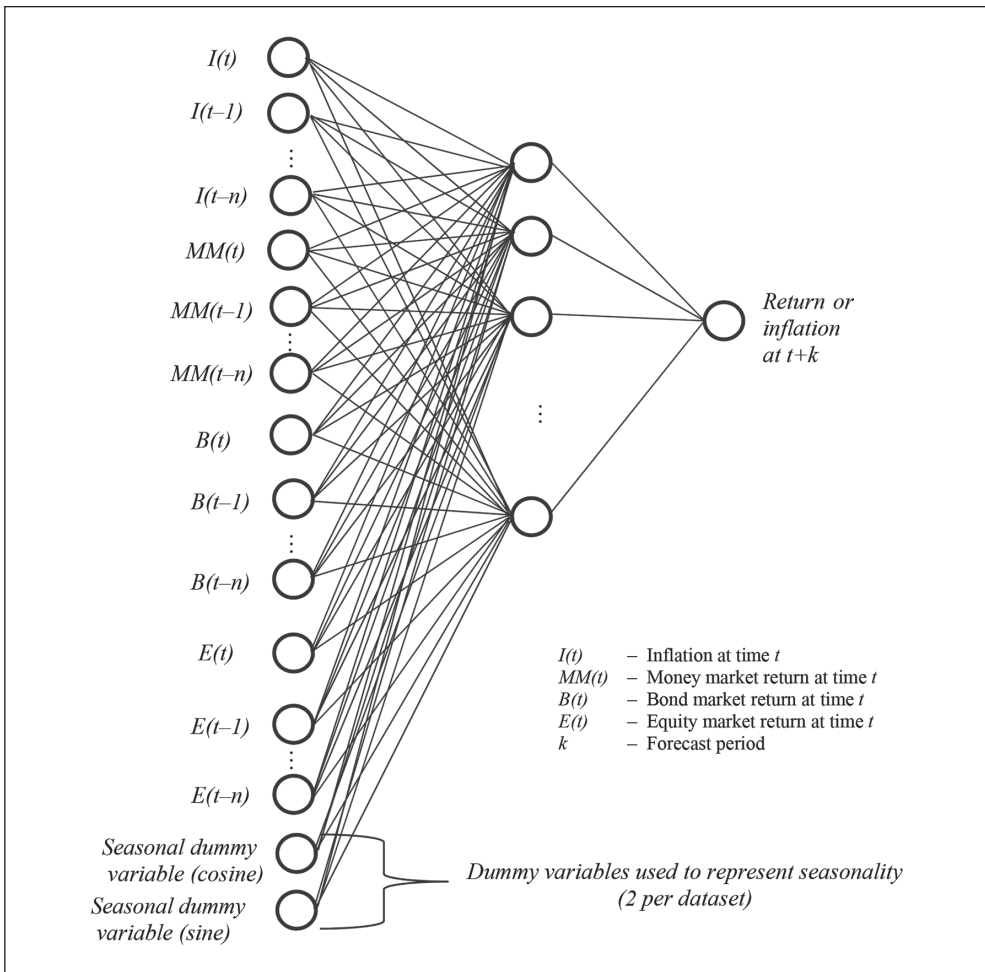


FIGURE 10. Design of returns and inflation forecasting, integrated ANN

argument holds with regard to the exponential effect. Three different learning rate increases are considered, i.e., 1.005, 1.05 and 1.1 (0.5%, 5% and 10% increase). These are chosen because small learning rates would ensure that a local minimum error is obtained; however, it would require significantly more training epochs. Likewise, the learning rate decreases are 0.2, 0.5 and 0.8 and correspond with an 80%, 50% and 20% decrease, respectively. Only the inflation forecasting is provided with seasonal explanatory variables in the isolated ANN; however, all the datasets of the integrated ANN are provided with seasonal explanatory variables.

6. MODELS

6.1 Artificial Neural Networks

The estimated parameters associated with the ANN for each application are presented in Table 4.

TABLE 4. Parameters of efficient ANNs

	Structure (input neurons—hidden neurons—output neurons)	Efficient epochs	Learning rate increase	Learning rate decrease
1-Step (isolated)				
Inflation	24–32–1	3500	0.50%	20%
Money market	6–32–1	100000	0.50%	20%
Bond market	6–32–1	2000	5.00%	20%
Equity market	6–16–1	49	5.00%	20%
3-Step (isolated)				
Inflation	12–32–1	4999	5.00%	20%
Money market	6–64–1	49999	0.50%	20%
Bond market	12–32–1	59	5.00%	20%
Equity market	6–32–1	50	5.00%	20%
1-Step (integrated)				
Inflation	24(6×4)–16–1	850	0.50%	20%
Money market	24(6×4)–16–1	10000	0.50%	20%
Bond market	12(3×4)–32–1	250	0.50%	20%
Equity market	12(3×4)–32–1	2200	5.00%	20%
3-Step (integrated)				
Inflation	48(12×4)–16–1	4000	0.50%	20%
Money market	12(3×4)–8–1	4200	0.50%	20%
Bond market	48(12×4)–8–1	50	5.00%	20%
Equity market	12(3×4)–16–1	50	5.00%	50%

6.2 Best-Fit Time Series Models

Table 5 indicates the types of ARIMA models that are used for each market, chosen according to the methodology described in section 3.4.

TABLE 5. Best-fit ARIMA models

Application	Model type
Inflation	Seasonal dummies + ARMA(2,1)
Money market	ARI(2,1)
Bond market	AR(1)
Equity market	Mean

7. RESULTS

7.1 Model Comparison

7.1.1 This section compares the accuracy of the different modelling approaches followed in this paper. First the approach used to compare the models is explained, followed by the analysis of results of the comparison.

7.1.2 The coefficient of determination (R^2 value) is used to compare the fit of the respective models. The R^2 value represents the proportion of the total variation of the observed data that is explained by the model and measures how well the forecast values compare to the observed values. The value of this measure exists in the interval $(-\infty; 1]$. The risk-return profile of returns is broadly as expected from the theory. The interpretation of the values of this measure is as follows: if $R^2=1$, the regression perfectly fits the data. If $0 \leq R^2 < 1$, the fit depends on the value of the measure. Values close to one indicate a good fit and values close to zero indicate a poor fit. It is also possible for $R^2 < 1$ as can be seen from the definition given in (10). A negative value indicates that the regression does not fit the data well and the model is a poor fit or even mis-specified. In this instance, the mean of the dataset will generally provide a better fit than the model.

$$R^2 = 1 - \frac{SSE}{SST} \quad (10)$$

where $SSE = \sum_{i=1}^n (\hat{x}_i - x_i)^2$ and $SST = \sum_{i=1}^n (x_i - \bar{x}_i)^2$. In this instance, n is the number of data points, \hat{x}_i is the forecast value for data point i , and x_i is the actual result of data point i .

7.1.3 Conventionally, the adjusted R^2 value is used in the evaluation of a model, as it accounts for the parsimony of the model. This is done by incorporating the number of parameters required by the model into the calculation of the measure. In most cases, this provides a good balance between the accuracy and the economy of a model. In the case of ANNs, the number of parameters does not necessarily increase as the system becomes increasingly complex. Therefore, using the adjusted R^2 measure will skew the results, as only

the ARIMA models will be kept parsimonious. For this reason, the R^2 measure is used in this study.

7.1.4 An R^2 threshold of 0.3 is used for illustrative purposes to evaluate the effectiveness of the different models on the basis that an R^2 value of 0.3 implies that a model explains a fair amount of variation in the data. However, more research should be done to determine an optimal threshold value. Figure 11 shows the plotted R^2 value of each application, including the threshold.

7.1.5 Figure 11 indicates that, for all the models of inflation, the R^2 values are below the threshold for both the one- and three-month forecasting period. The integrated ANN and the ARIMA model explain approximately 25% of the variation within the inflation data when a single-month forecast period is considered. The isolated ANN explains slightly over 15%. This finding indicates that all the models explain a significant amount of variation within the inflation data for a single-month forecast period. The significantly higher R^2 value for the ARIMA model for forecasting inflation three months ahead can be explained by the seasonal trend being better captured. The ANNs and ARIMA model explain a fair amount of variation in the inflation data when a single-month forecast period is considered. As the forecast period increases, the R^2 value decreases considerably for both the ANNs and the ARIMA models.

7.1.6 Considering the money market, the R^2 values of all the models are above the threshold for both the one- and the three-month forecasts. The models explain over 80% of the variation within the money market, which suggests that this market can be modelled

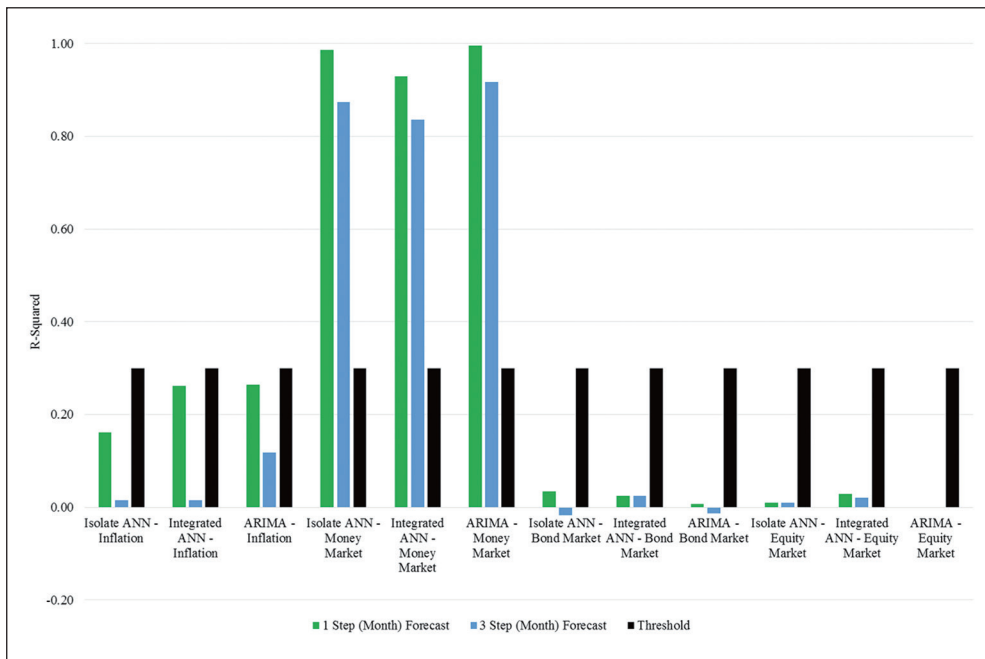


FIGURE 11. R^2 values of forecasting models—testing dataset

as a strict time series over the short term. The high predictability of this market suggests that there is little random variation within the data, which was expected, as the money market is stable over the short term. Furthermore, this finding suggests that the trends within the money market do not change regularly and are not volatile. All the models constructed to forecast the money market, for one and three months into the future, can be used in practice.

7.1.7 The R^2 values for the models associated with the bond market are significantly smaller than the threshold value. Slightly negative R^2 values are observed for the isolated ANN and the ARIMA model for the three-month forecasts, which indicates that these models might not outperform the mean of the data. All the models explain less than 5% of the variation within the bond market. This finding indicates that none of the models are effective at forecasting the returns on the bond market, either for one or for three months into the future. The accuracy of the models could possibly be improved by incorporating external explanatory variables.

7.1.8 The equity market models explain less than 5% of the variation within this market. This poor performance is expected, as the equity market is highly volatile and heavily dependent on external factors, which are not included in the models. Therefore, incorporating these factors could possibly increase the accuracy of the forecasts.

7.1.9 Overall, the ANNs and the time series models that are used to forecast the money market performed well for both the one- and the three-month forecast periods and could be used in practice. The models used to forecast inflation over a one-month period explain a fair amount of variation within the inflation data; however, their performance is poor when the forecast period is extended to three months. The models used to forecast the bond and equity market returns performed poorly over both the one- and three-month forecast periods.

7.2 Relationship between Error and Forecast Period

7.2.1 The RMSE of the forecasts is used to evaluate the relationship between the forecasting accuracy and the period of the models. The RMSE exists in the interval $[0; \infty)$ and is given by

$$RMSE = \sqrt{\frac{1}{n} \sum_{t=1}^n \varepsilon_t^2} \quad (11)$$

where $\varepsilon_t = (\hat{x}_t - x_t)$, \hat{x}_t is the forecast value at time t , X_t is the actual value at time t , and n is the number of forecasts generated. The RMSE is interpreted as follows: if $RMSE = 0$, the observed and forecast variables are identical. However, if $RMSE > 0$, there is a certain level of discrepancy between the forecast and the observed values of the variable. In particular, if $RMSE$ is much greater than zero, there is a large error between the forecast and the observed values of the variable. The size of the error is an indicator of the reliability of the model.

7.2.2 Figure 12 shows the RMSE plotted for each model for each application. As the figure shows, the RMSEs associated with the inflation models increase as the forecast period expands from one to three months. The greatest increase in RMSE occurred with

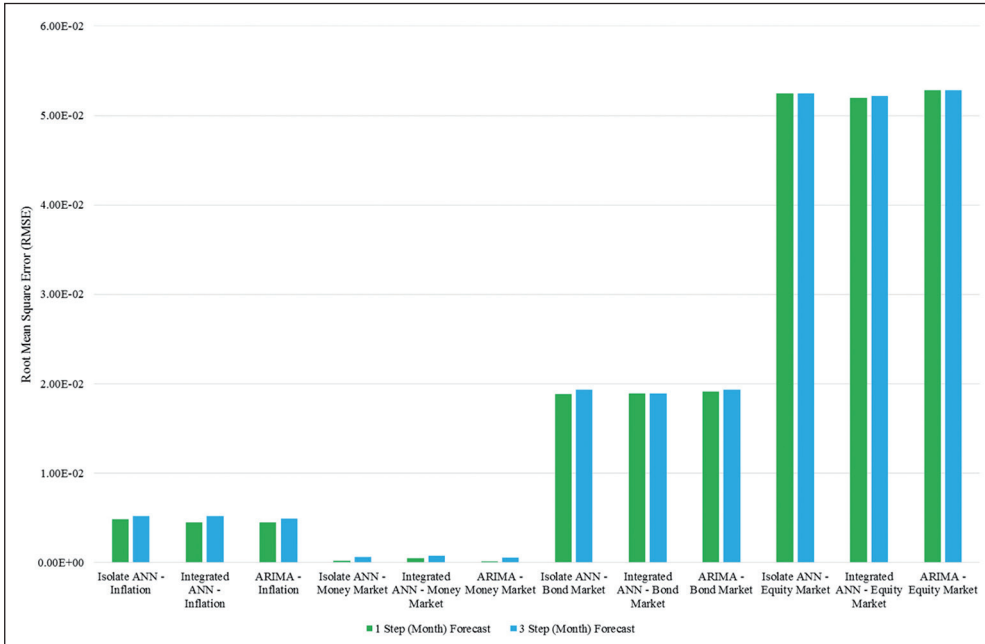


FIGURE 12. RMSE of forecasting models—testing dataset

respect to the integrated ANN. This finding could be explained by the ANN capturing random variation, because of overfitting, when the forecast period is extended. The performance of the ARIMA model is the most efficient for the extended forecast period, as is proved by this model having the smallest RMSE over the three-month period. This implies that the compounding of errors in the forecasts associated with the ARIMA model (over one- or three-month periods) did not lead to inferior results compared with the ANNs. For ANNs such error compounding does not occur as each model is trained to forecast either one or three months in advance.

7.2.3 As an illustration, to assess the effect of the compounding of errors for the time series models, a longer forecasting period is considered for the money market. When considering a twelve-month forecasting period for the money market, the accuracy of the ANNs exceeded that of the time series models. The increase in accuracy is indicated by the reduced error over the testing set of data given in Table 6.

7.2.4 A general and significant increase in RMSE is found for all the money market models for the extended forecast period. The isolated ANN has the greatest and the integrated ANN the smallest increase. The integrated ANN captured random variation when both periods are considered. The ARIMA time series model showed a significant increase in RMSE as the forecast period increased; however, this model has the smallest RMSE over both the one- and the three-month forecast periods.

7.2.5 When the bond market is considered for an extended forecast period, the RMSE of the forecasts increases for both the isolated ANN and the ARIMA model. The

increase is lower than expected, and could be explained by the poor forecasting accuracy of the models associated with the bond market. No increase occurs when the integrated ANN is considered, which is unexpected. This anomaly could possibly be explained by a combination of poor forecasting accuracy and random variation.

7.2.6 With regard to the equity market, there is no increase in the RMSE of the forecasts for either the isolated ANN or the ARIMA model. Furthermore, the increase in RMSE for the integrated ANN is slight. Again, this result could be explained by the poor forecasting accuracy of the models over both the one- and the three-month periods.

7.2.7 In general, the error of each model increases as the forecast period is extended from one to three months. This was to be expected, as forecasts over the near future are generally more predictable than the distant future.

TABLE 6. RMSE for twelve-step forecasts of the money market

Twelve-step forecasts—money market		
	Training set	Testing set
ARIMA model	0.003163	0.002126
Isolated ANN	0.00434	0.001834
Integrated ANN	0.004725	0.001967

7.3 Inter-Market Relationships

With regard to the one-month forecasts, the integrated ANN only generates more accurate forecasts for the rate of inflation and the equity market. These observations imply that the relationships between the markets and inflation have a smaller effect on forecasting accuracy over a one- and three-month period than expected. Regarding the three-month forecasts, increased forecasting accuracy is seen in the integrated ANNs when the bond and equity markets are considered. However, the relationships between the markets and the rate of inflation do not add value to the forecasts in the instances of inflation and the money market. The fact that the integrated ANNs do not produce significantly more accurate forecasts, suggests that the modelling of dependencies between these markets may not lead to significantly improved results.

7.4 Artificial Neural Networks compared with Best-Fit Time Series Models

7.4.1 As regards inflation, none of the isolated ANN, integrated ANN or ARIMA models appear superior for one-month inflation forecasts. The RMSE of the integrated ANN and ARIMA model are found to be similar, and the isolated ANN achieves the greatest RMSE. Furthermore, the residuals of the isolated ANN contain the greatest volatility, which is not a desirable result. The ARIMA model has the smallest volatility within its residuals. No significant difference is found between the performances of the models when considering the three-month forecast period. As CPI data from 2011 to 2013 are freely available, this data was used as a validation set to test the isolated ANN and the ARIMA model.

7.4.2 Figure 13 depicts the plot of the comparison intervals (as defined in ¶3.5.2) of the isolated ANN, integrated ANN, and the ARIMA model that are used to forecast the money market one month into the future. The horizontal line on the chart links the RMSE of each model, and the vertical line illustrates the comparison interval for each model.

7.4.3 With respect to the money market, Figure 13 indicates that the ARIMA model has both the smallest RMSE and volatility within its residuals when forecasting one month ahead. The volatility of the residuals of the ARIMA model is significantly smaller than that of the isolated or the integrated ANNs. The integrated ANN has both the greatest RMSE and volatility within its residuals, which suggests that the integrated ANN captures the limited random variation present in the data. The isolated ANN has an intermediate RMSE and volatility within its residuals. None of the models could be classified as superior; however, the ARIMA model has the smallest RMSE and smallest volatility within its residuals. The ARIMA model is therefore considered the preferred model for forecasting the money market one month in advance.

7.4.4 Figure 14 indicates that the ARIMA model, again, achieves the smallest RMSE and the smallest volatility within its residuals when forecasting the money market three months in advance. The RMSE of the forecasts and the volatility within their residuals does increase from the one-month forecasts, as expected. The RMSE of the forecasts generated by the isolated ANN is not significantly greater than those generated by the ARIMA model. Again, the integrated ANN has both the greatest RMSE and greatest volatility within its residuals. The ARIMA model is preferred to the ANNs in this instance, similarly to the one-month forecast. As regards a twelve-month forecast period, the isolated and the integrated ANNs have a smaller RMSE than the ARIMA model. Furthermore, the residuals of the forecasts generated by the ANNs are less volatile than those associated with the ARIMA model. This suggests that, as the forecast period is extended, the compounding of the errors associated with the forecasts from the ARIMA model has a significant effect on the accuracy of the model. Furthermore, as all the comparison intervals overlap, it is not possible to

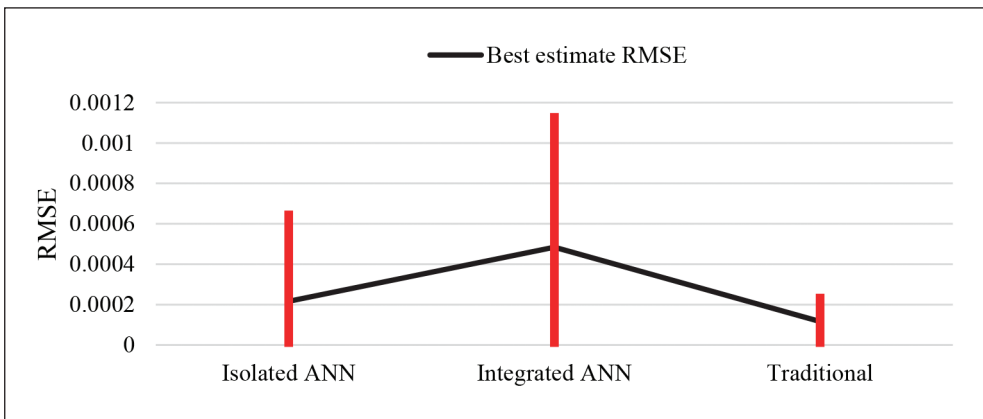


FIGURE 13. Best estimate and comparison interval for money market models— one step/month forecasts

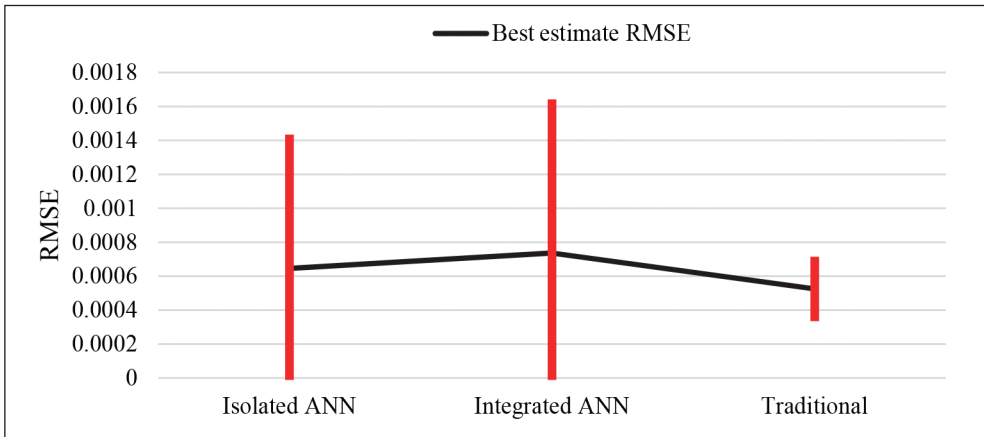


FIGURE 14. Best estimate and comparison interval for money market models—three step/month forecasts

conclude that any model performs significantly better than the others. However, the isolated ANN approach is preferred, seeing that it yields the smallest RMSE and the least volatility within its residuals.

7.4.5 For both the bond and the equity markets, no significant differences are found between the performance of the isolated ANN, the integrated ANN and the ARIMA models under one- and three-month forecast periods.

7.4.6 In general, no model proved significantly superior to the others. In particular, the integrated ANNs did not perform significantly better than isolated ANNs as might be intuitively expected. Forecasting accuracy is slightly influenced by the relationship between the various markets and inflation. These effects are relatively small, implying that the relationships between the markets and inflation are not necessarily a significant driver of these indices. With regard to forecasting the money market over a period of one and three months, the ARIMA model is preferred over the ANNs because of the smaller RMSE and volatility within the residuals of this model. However, the opposite is seen when forecasting the money market twelve months ahead, because the forecasting error associated with the ARIMA model began to compound as the forecast period was extended. The result was reduced accuracy and increased volatility within the residuals of the ARIMA models for a forecast period of twelve months.

7.4.7 From the investigations performed, it can be concluded that both the isolated and the integrated ANNs perform as well as the fitted ARIMA models when forecasting the return on the money, bond and equity markets, as well as the rate of inflation.

7.5 Training- and Testing-Set Anomaly

It is noted that the RMSE over the testing set is smaller than the RMSE over the training set. This anomaly occurs for both the time series models and the ANNs. The observed anomaly could be explained by the increased variation contained in the specific training set.

If this variation are removed from the dataset, the error over the training set would be smaller than it is over the testing set. This finding is in agreement with expectation.

7.6 Hybrid Models

7.6.1 A slight difference is found between the RMSE and the volatility within the residuals of the models when forecasting inflation one month into the future. The smallest RMSE is achieved by the hybrid model with an isolated ANN. This model also has slightly less volatility within its residuals than do the other models. The hybrid model with an integrated ANN has the greatest RMSE and volatility within its residuals, which indicates that the hybrid model captures random variation within the data. No model performed significantly better than the others in this application. This trend extended to all the market models when generating one-month forecasts.

7.6.2 When forecasting inflation three months ahead, the models have a similar RMSE. The volatility within the residuals of the ARIMA model was the greatest. The hybrid models have similar volatility in their residuals, which suggests that the hybrid models result in less volatile forecasts when forecasting inflation three months in advance. For this analysis, it is not possible to conclude that any model is statistically more accurate than any other at forecasting inflation three months in advance.

7.6.3 Figure 15 shows the plot of the comparison intervals of the hybrid model consisting of the isolated ANN, the hybrid model consisting of the integrated ANN, and the ARIMA model that are used to forecast the money market three months into the future. The horizontal line on the chart links the RMSE of each model, and the vertical line captures the comparison interval for each model.

7.6.4 The comparison intervals of the three-month forecasts of the money market (Figure 15) indicate that the RMSE of all three models are similar. The hybrid model with

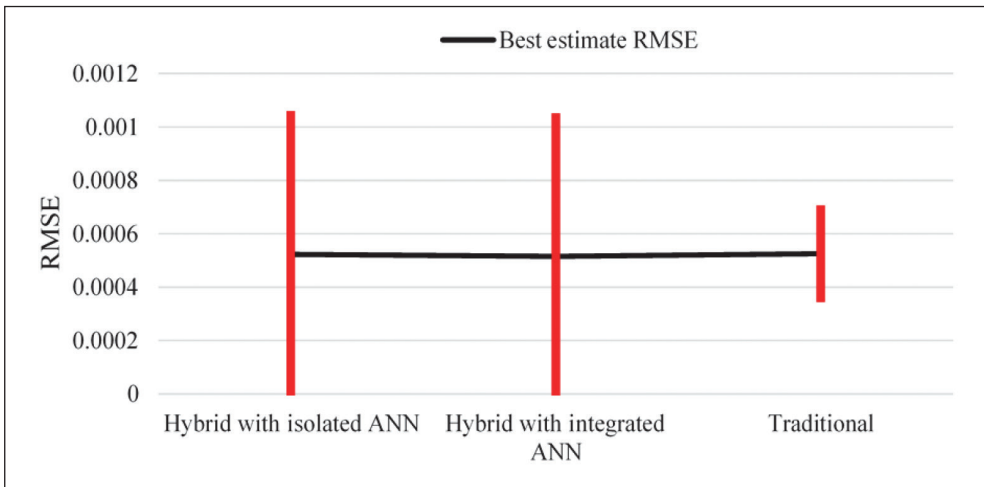


FIGURE 15. Best estimate and comparison interval of RMSE for hybrid money market models—three step/month forecasts

an integrated ANN achieves a slightly lower RMSE than the other models. However, the residuals from the ARIMA model have the least volatility of all the models considered.

7.6.5 No model type could be classified as statistically superior; however, the hybrid model is preferred when forecasting inflation because of the reduction in volatility within the residuals of the forecasts generated by this model. It can therefore be concluded that ANNs are likely to add value to the time series models through the hybrid models when forecasting inflation. The ARIMA model is preferred when forecasting the return on the money market, as the volatility within its residuals is the least.

7.7 Cross-Validating Results

7.7.1 Figure 16 shows the plot of the confidence intervals of the isolated ANN, integrated ANN, and the ARIMA model that are used to forecast inflation one month into the future. The horizontal line on the chart connects the RMSE of each model, and the vertical line illustrates the comparison interval for each model.

7.7.2 Figure 16 indicates that the isolated ANN generates forecasts that, on average, are more accurate than those generated by the other models. The forecasts of the integrated ANN produces the greatest RMSE on average over the five datasets considered. It is concluded that, as all the confidence intervals overlap, no model could be regarded as statistically superior with regard to the forecasting of the rate of inflation one month in advance.

7.7.3 The performance of the integrated ANN was expected to be superior to that of the other models, as it allowed interaction between the markets and inflation. However, these interactions are less influential than expected and cause additional random variation to be captured by the ANN. On average, the isolated ANN performs better than the other models. The variation of the RMSE is similar for all three models. These conclusions are based on the

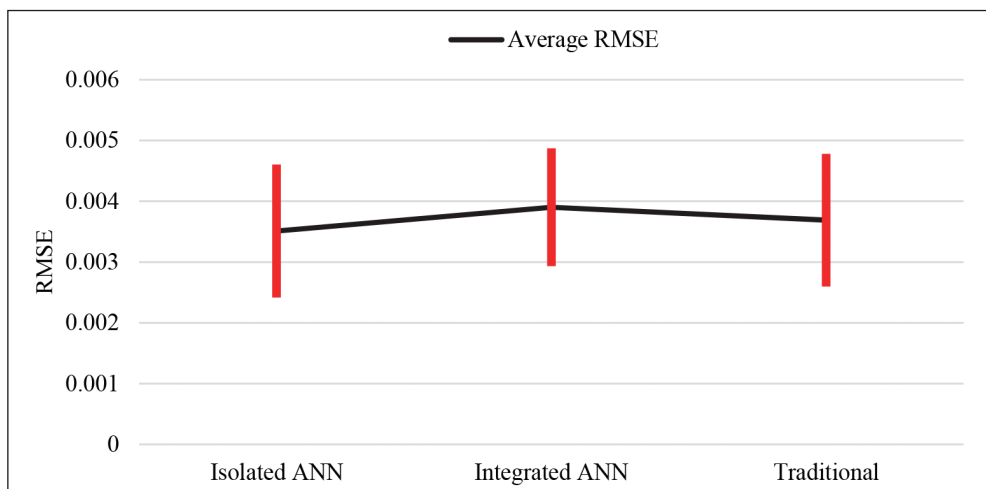


FIGURE 16. Confidence intervals of RMSEs of models over five different datasets—one step/month inflation forecast models

results from five different training and testing datasets; however, for the conclusions to have more statistical credibility, it is necessary to consider more than five sets of data.

8. CONCLUSIONS

8.1 An extensive methodology is proposed in this paper to fit a FF MLP ANN for financial applications. In order to constrain the scope of the study to a manageable extent the authors selected a FF MLP ANN as the focus of the study. This guides the designer of the ANN to choose the structure and parameters of the network that should provide the best design amongst a set of candidate designs and a range of parameters. As case studies, custom ANN structures are determined for the money, bond and equity market as well as inflation.

8.2 A FF MLP ANN structure is chosen for this study. This necessitates modelling decisions regarding the number of input (lagged observations) neurons, the number of hidden neurons, the activation function and the weight initialisation. Weights are initialised randomly using Monte Carlo methods. The number of input and hidden neurons are chosen using a heuristic search. The authors chose a hyperbolic tangent activation function as it has a continuous range $(-1, 1)$ as required by this application. Modelling decisions pertaining to the structure of the ANN are worthy of further investigation.

8.3 The learning algorithm parameters that are determined include the learning rate increase and the learning rate decrease.

8.4 Considering a one-month forecast period, only the forecasts generated for inflation and the money market could be considered for practical use. For a three-month forecast period, only the forecasts generated for the money market could be used in practice. The bond and equity market forecasts generated by all the models are poor.

8.5 In general, the forecasting accuracy associated with each market decreases as the forecast period expands. As seen in the results for the bond and equity markets, this can be distorted by random variation. The integrated ANN resulted in a slightly lower RMSE and slightly lower volatility within the forecast residuals than did the isolated ANNs when forecasting inflation and the return on the equity market. The opposite is true for the money market and the bond market. The difference in performance of these models is slight and cannot be generalised. As the integrated ANNs do not perform significantly better than the other models, the inter-correlation between market movements and inflation are less pronounced than might be expected.

8.6 It is not possible to conclude that the performance of any model is significantly superior to that of the others. However, with regard to the money market, the ARIMA model is preferred for both the one- and three-month forecasts. This finding suggests that ANNs perform as well as the ARIMA time series models considered when forecasting inflation

and the return on the bond and equity markets in South Africa over a period of one and three months. ARIMA models are preferred, on the other hand, when forecasting short-term returns on the money market. The forecast accuracy of all the models diminishes over a longer forecast period. This effect appears to be less pronounced for the ANN model. When the twelve-month forecasting period is considered, the ANNs have a smaller error and are preferred. As the forecasting period increases, ANNs outperform their time series counterparts when considering the money market.

8.7 The general performance of ANNs is not significantly different from the corresponding performance of the time series models considered, when forecasting inflation over a single month. This is suggested by the results obtained when the models are applied to different subsets of inflation data. Furthermore, ANNs add value to time series models through the hybrid models when forecasting inflation.

9. AREAS OF FURTHER RESEARCH

The methodology presented in this paper indicates that scope for further investigation exists.

- ANNs with a single hidden layer are used in this study; however, adding additional hidden layers could be considered. Expanding the number of layers could increase the accuracy and reduce the complexity of the ANNs.
- The learning algorithm could be improved. Introducing more complex second-order learning algorithms could lead to reduced training time.
- The size of the training and testing datasets could be altered to determine the optimal size to balance both the volume of data and the statistical credibility of the results.
- A validation dataset, as used for inflation, could be introduced for the other applications. This set could consist of the returns on the relevant markets from 2011 to 2013.
- The learning rates and weights are initialised using a normal distribution. Other distributions could be considered to determine which is optimal for the initialisation process.
- Different ANN types and structures could be considered to identify alternatives providing a better balance of complexity and accuracy.
- External explanatory variables could be introduced into the models to improve the forecast accuracy.
- Parameter and structure searches could be performed using local-gradient based or global optimisation algorithms.
- Different activation functions could be tested to gauge their effect on model performance.

ACKNOWLEDGEMENT

This research was undertaken, in part, thanks to funding from the Barclays Africa Chair in Actuarial Science.

REFERENCES

- Abhishek, K, Khairwa, A, Pratap, T & Prakash, S (2012). A stock market prediction model using Artificial Neural Network. *Computing Communication & Networking Technologies (ICCCNT)*, Third International Conference, 1–5
- Allard, R & Faubert, J (2004). Neural networks: different problems require different learning rate adaptive methods. *Proceedings of SPIE 5298 (International Society for Optical Engineering)*, Image Processing: Algorithms and Systems III, 516–27. DOI:10.1117/12.527094
- Bouqata, B, Bensaid, A & Palliam, R (1999). Recurrent artificial neural networks for forecasting of forward interest rates. *International Joint Conference on Neural Networks*, Vol. 6, 3894–97
- Braun, H & Lai, LL (2005). A neural network linking process for insurance claims. *Proceedings of 2005 International Conference on Machine Learning and Cybernetics*, 399–404
- Fei, W & Zhigang, S (2008). Research on the credit risk evaluation and forecast of housing mortgage loans. *Intelligent Information Technology Application Workshops (IITAW) 2008*, 940–43
- Firer, C & McLeod, H (1999). Equities, bonds, cash and inflation: historical performance in South Africa. *Investment Analysts Journal*, **28**(50), 7–28
- Firer, C & Staunton, M (2002). 102 Years of South African financial market history. *Investment Analysts Journal*, **31**(56), 57–65
- Huang, G-B, Zhu, Q-Y & Siew, C-K (2006). Extreme learning machine: theory and applications. *Neurocomputing*, **70**, 489–501
- Huang, Y (2009). Advances in artificial neural networks – methodological development and application. *Algorithms*, **2**(3), 973–1007. DOI:10.3390/algorithm2030973
- Kohzadi, N, Boyd, MS, Kermanshahi, B & Kaastra, I (1996). A comparison of artificial neural network and time series models for forecasting commodity prices. *Neurocomputing*, **10**, 169–181
- Korol, T (2013). Early warning models against bankruptcy risk for Central European and Latin American enterprises. *Economic Modelling*, **31**, 22–30
- Lacher, RC, Coats, PK, Sharma, SC & Fant, LF (1995). A neural network for classifying the financial health of a firm. *European Journal of Operational Research*, **85**, 53–65
- LeCun Y, Bengio, Y & Hinton, G (2015). Deep learning. *Nature*, **521**, 436–44
- O'Connor, N & Madden, MG (2006). A neural network approach to predicting stock exchange movements using external factors. *Knowledge-Based Systems*, **19**, 371–7
- Riedmiller, M & Braun, H (1993) A direct adaptive method for faster backpropagation learning: the RPROP algorithm. *Neural Networks, IEEE International Conference*, Vol. 1, 586–91
- Russell, S & Norvig, P (2010). *Artificial Intelligence – A Modern Approach*, 3rd edition, Pearson
- Saha, SP (2009) On small sample prediction of financial crisis. *Seventh International Conference on Advances in Pattern Recognition (ICAPR) 2009*, 43–46
- Sahin, Y & Duman, E (2011). Detecting credit card fraud by ANN and logistic regression. *International Symposium on Innovations in Intelligent Systems and Applications (INISTA) 2011*, 315–19
- Schmidt, A (2000). Biological neural networks. www.teco.edu/~albrecht/neuro/html/node7.html
- Tan, CNW (1999). An artificial neural networks primer with financial applications examples in financial distress predictions and foreign exchange hybrid trading system. *School of Information Technology, Bond University, Gold Coast*

- Tsai, C-F & Wu, J-W (2008). Using neural network ensembles for bankruptcy prediction and credit scoring. *Expert Systems with Applications*, 34, 2639–49
- Wilamowski, BM (2009). Neural network architectures and learning algorithms. *Journal of Machine Learning Research* 12, 2493–537
- Zhang, G, Patuwo, EB & Hu, MY (1998). Forecasting with artificial neural networks: The state of the art. *International Journal of Forecasting*, 14, 35–62

APPENDIX

A.1 Results from Time Series Models

Summary of models							
Application	Model type	1 Step			3 Steps		
		RMSE training	RMSE testing	RMSE validation	RMSE training	RMSE testing	RMSE validation
Inflation	Seasonal dummies + ARMA(2,1)	0.004922	0.004492	0.003255	0.005032	0.004921	0.003223
Money market	ARI(2,1)	0.000305	0.000116		0.001073	0.000525	
Bond market	AR(1)	0.02566	0.01906		0.02586	0.01925	
Equity market	Mean	0.0651	0.05278		0.06522	0.05278	

A.2 Results from Artificial Neural Networks

Summary of the results of the experiment—errors			
	RMSE training	RMSE testing	RMSE validation
1-Step (isolated)			
Inflation	0.0048	0.0048	0.0031
Money	0.000474	0.000217	
Bond	0.0233	0.0188	
Equity	0.0652	0.0525	
3-Step (isolated)			
Inflation	0.0049	0.0052	0.0033
Money	0.0011	0.000645	
Bond	0.0260	0.0193	
Equity	0.0647	0.0525	
1-Step (integrated)			
Inflation	0.0045	0.0045	
Money	0.000496	0.000484	
Bond	0.0252	0.0189	
Equity	0.0514	0.0520	
3-Step (integrated)			
Inflation	0.0047	0.0052	
Money	0.000908	0.000736	
Bond	0.0248	0.0189	
Equity	0.0655	0.0522	

A.3 Results of Hybrid Models

Hybrid models—errors		
RMSE		
Inflation	Training set	Testing set
1-Step		
Isolated	0.004926	0.004436
Integrated	0.005145	0.00469
3-Step		
Isolated	0.005316	0.004807
Integrated	0.005194	0.004916
Money market		
1-Step		
Isolated	0.000301	0.000117
Integrated	0.000313	0.000116
3-Step		
Isolated	0.001077	0.000523
Integrated	0.001146	0.000516

A.4 Results of Cross-Validation

RMSE over testing set for five independent datasets			
Datasets	ARIMA model	Isolated ANN	Integrated ANN
1	5.44E-03	5.26E-03	5.02E-03
2	3.47E-03	3.40E-03	3.16E-03
3	3.58E-03	3.14E-03	4.48E-03
4	2.94E-03	3.16E-03	3.99E-03
5	3.01E-03	2.59E-03	2.86E-03