ASIAN ACADEMY *of* MANAGEMENT JOURNAL *of* ACCOUNTING *and* FINANCE

AAMJAF, Vol. 5, No. 2, 101–118, 2009

NONLINEAR PREDICTION OF THE STANDARD & POOR'S 500 AND THE HANG SENG INDEX UNDER A DYNAMIC INCREASING SAMPLE

Manish Kumar*

Department of Management Studies, IIT Madras, Chennai 600036 India

E-mail: manishkumar_iitm@yahoo.com.in

ABSTRACT

This study attempts to forecast the next day's returns of two time series in the Hang Seng Index (HSI) and Standard & Poor's (S&P) 500 indices using Artificial Neural Networks (ANN) with past returns as input variables. Results from ANN are compared with those from the autoregressive integrated moving average (ARIMA) model. This study uses a longer time period than ARIMA (i.e., daily data of 80 and 35 years for the S&P 500 and HSI, respectively) to develop and test the models. The two competing models are rigorously evaluated in terms of widely-used penalty-based criteria, such as directional accuracy, as well as in terms of trading performance criteria like annualised return, the Sharpe ratio and annualised volatility via a simple trading strategy. Moreover, the robustness of the two models is tested for 36 test periods. Empirical results show that ANN works better than ARIMA and delivers consistent results across the periods tested. These results support ANN's robustness and its use in formulating a strategy for trading in the S&P 500 and HSI time series.

Keywords: ARIMA, artificial neural network, forecasting, stock market, time series analysis

INTRODUCTION

Financial time series analysis is considered one of the most difficult tasks because time series are inherently noisy, nonstationary and deterministically chaotic (Yaser & Atiya, 1996). Moreover, such series are influenced by numerous factors, including human error, political contexts, economic situations and competition. Thus, modelling and forecasting the movement of such series is quite difficult. However, accurate predictions of various financial time series is of

* The author is currently attached at IREVNA, Division of CRISIL, Chennai 600017 India.

paramount importance to hedge against potential market risk and create new profit-making opportunities (Manish & Thenmozhi, 2007). Therefore, developing forecasting models for financial time series has practical, as well as theoretical, importance. The practical aspect involves predictive power, which provides useful information to a variety of actors, including investors, regarding asset allocation decisions, firms in risk management and regulators in policymaking. Theoretically, predictability about returns can lead to important contradictions to the efficient market hypothesis.

Thus, a large number of academics, practitioners and regulators use different techniques to forecast financial time series. This forecast is commonly done using various statistical models (Granger & Newbold, 1986). These models can be classified as univariate and multivariate models. One of the most commonly-used univariate time series models is the autoregressive integrated moving average (ARIMA) model; the most common of the multivariate framework is the regression model. Much of early forecasting was based on these conventional models (Richard & Wichem, 1998).

A key constraint of the ARIMA and multiple regression models is the pre-assumed linear structure of the model. Financial time series are nonlinear in nature; thus, the linear model estimates for the financial time series problem are not always satisfactory (Refenes et al. (1994)). In the last decade, however, Artificial Intelligence (AI) techniques such as Artificial Neural Networks (ANN) have been used successfully in a number of time series forecasting applications (Azoff, 1994, Zhang et al.,1998, Yu et al. 2005).

ANN offers various advantages. It makes very few assumptions relative to the normality assumptions commonly found in statistical methods. ANN can perform predictions after learning the underlying relationship between the input variables and the output. It is analogous to nonparametric, nonlinear regression models. Hornik, Stinchcombe and White, (1989) argued that ANN can approximate a rather general family of nonlinear functions, as it has the so-called universal approximation property. As such, a number of studies have focused on the application of ANNs to stock market predictions (Ahmadi, 1990, Kimoto, Asakawa and Takcoka, 1990), Trippi et al. (1992), Choi et al. (1995), Donaldson and Kamstra, 1996, Zirilli, 1997, Thenmozhi, 2001, Kim, 2003, Zhang, 2003, Manish and Thenmozhi 2007, Tugba and Casey 2005, Huang et al. 2005, Manish and Thenmozhi 2007). However, in most studies, statistical models such as ARIMA have been used as a benchmark to evaluate ANN's performance (Hwarng & Ang, 2002).

The application of ANN in finance and economics has gained momentum in tandem with the study of models of volatility. The Autoregressive Conditional

Heteroskedasticity/Generalised Autoregressive Conditional Heteroskedasticity (ARCH)/GARCH) family of models has been used to model the heteroscedasticity of financial time series. When the nonlinearity in a series is due to GARCH, ANN cannot make accurate forecasting. But if the nonlinearity is not due to GARCH, then ANN does help in exploiting the extra structure, thereby producing precise forecasts. This issue has received little attention in the literature.

There are other vital problems with earlier studies. In most studies, the degree of accuracy and the acceptability of forecasting models were measured by the estimate's deviations from observed values and have not considered turning-point forecast capability using sign and direction tests. Leung et al. (2000) suggested that depending on the investor trading strategies, forecasting methods based on minimising forecast error may not be adequate to meet objectives. Thus, competing models must be evaluated not only in terms of Mean Square Error (MSE) and Mean Absolute Error (MAE) but also in terms of sign and direction.

Most studies also do not evaluate their models based on trading performance. The forecast error may have been minimised during model estimation, but a model with a small forecast error may not be as profitable as a model selected using financial criteria, such as a risk-adjusted measure of return (Leung et al. (2000)). Therefore, an evaluation of models using financial criteria based on a trading experiment may be more appropriate.

Though previous studies focus on out-of-sample performance, most studies arbitrarily split the available data into a training (i.e., in-sample) set for model construction and a test (i.e., out-of-sample) set for model validation, which leads to two connected problems. First, it may introduce bias into model selection and evaluation. Second, it ignores the effect of sample size. The differences in model performance are likely to be a result of variations in the time frame and the number of observations used. Last but not the least, in almost all studies, the number of observations for training and the amount of test data were very low.

The present study overcomes the drawbacks identified in the earlier studies by examining the applicability of ARIMA and ANN for predicting the daily returns of the S&P 500 in the US and the HSI in Hong Kong. Major contributors of this study is to tackle the problem of sampling variation, by employing a cross-validation method to examine the out-of-sample performance of the two time series models.

The two competing models are rigorously compared using two approaches. First, the study examines the out-of-sample forecasts generated by different competing models employing the statistical criteria, that is, the proportion of times the signs of returns are correctly forecasted (signs, success or hit ratio (HR)). Second, the two competing models are also examined in terms of trading performance and economic criteria through a trading experiment. In the last step, we eradicate the heteroscedasticity effect from the two series and again perform the forecasting exercise using ARIMA and ANN models.

The rest of this paper is organised as follows. Section 2 describes the daily closing price of the S&P 500 and HSI indices and the conceptualisation of the ARIMA and ANN models. In Section 3, empirical results are presented. Finally, Section 4 concludes the paper with some discussion of future research.

DATA AND METHODS

The data set comprises the daily closing price of the S&P 500 and HSI indices. The HSI series spans 23 January 1975 to 31 December 2008, totalling 8,380 trading days. Meanwhile, the time period for the S&P 500 is 6 November 1928 to 31 December 2008, totalling 20,132 trading days. To overcome the problem of nonstationarity, the two series are transformed into rates of return. Daily returns are continuously compounded for rate of return, computed as the first difference of the natural logarithm of the daily index levels.

Dynamic Increasing Methodologies for Estimation and Prediction

To see how forecast performance is changing according to the choice of forecasting sample periods is not only empirically interesting but also a meaningful trial to confirm the robustness of the empirical results. To address the problems in sampling variation, this study uses a 36-period validation set for examining the out-of-sample performance of the ARIMA and ANN models. In particular, our study focuses on ANN robustness with respect to sampling variation.

The S&P 500 daily returns series is split into two periods; 6 November 1928 to 31 December 2005 is used for model estimation and classified as an in-sample period, and 1 January 2006 to 31 December 2008 is reserved for out-of-sample forecasting and evaluation. The out-of-sample period is further divided into 36 sub-samples, each a month long. A dynamically-increasing method is adopted to evaluate ARIMA and ANN performance on out-of-sample data. For instance, when the 6 November 1928 to 31 December 2005 period is used as an in-sample period, 1 January 2006 to 31 January 2006 is used as the out-of-sample period. For the in-sample period of 6 November 1928 to 31 January 2006, the corresponding out-of-sample period is 1 February 2006 to 28 February 2006. Similarly, for the last out-of-sample period of 1 December 2008 to 31 December

2008, the corresponding in-sample period is 6 November 1928 to 30 November 2008.

The dynamic increasing method used here is a recursive approach. However, in various empirical studies, the rolling approach is most common, as it works better than the recursive approach. However, Corradi and Swanson (2005 and 2006) suggest that if the in-sample portion of the data used for estimation is much larger than the out-of-sample portion of the data used for predictive accuracy testing and for model evaluation in general, then the contribution of the parameter estimation error is asymptotically negligible in the case of both the recursive and rolling estimation schemes. In this present study, we use a much larger data set and so have elected to use the recursive approach. Another advantage of using the recursive approach is that it makes use of all available data to estimate parameters (Stock & Watson, 2003).

Similarly, the HSI daily returns series is split into two periods; 23 January 1975 to 31 December 2005 is used for model estimation and classified as an in-sample period, and 1 January 2006 to 31 December 2008 is reserved for out-of-sample forecasting and evaluation. The entire out-of-sample period is further divided into 36 sub-samples, with each sub-sample lasting a month. A dynamically increasing method is adopted to evaluate the performance of the ARIMA and ANN models on out-of-sample data. For instance, if 1 January 1974 to 31 December 2005 is used as an in-sample period, 1 January 2006 to 31 January 2006 is used as an out-of-sample period. For the in-sample period 1 of January 1973 to 31 January 2006, the corresponding out-of-sample period is 1 February 2006 to 28 February 2006. Similarly, for the last out-of-sample period of 1 December 2008 to 31 December 2008, the corresponding in-sample period is 1 January 1974 to 30 November 2008. Hence, for every out-of-sample period of the S&P 500 and HSI returns, ARIMA and ANN performance is evaluated using HR, annualised returns, Sharpe ratio, maximum drawdown, annualised volatility and other data characteristics.

The ARIMA Model

Popularly known as the Box-Jenkins (BJ) method but technically known as the ARIMA model, the ARIMA takes the form:

$$Y_t = a_0 + \sum_{i=1}^{p} \alpha_i Y_{t-i} + \sum_{i=0}^{q} \beta_i e_{t-i} + \xi_t \tag{1}$$

where Y_t is the time series, and ξ_t is an uncorrelated random error term with zero mean and constant variance. a_0 is a constant term.

The correlogram, which is simply the plots of Autocorrelation Functions (ACFs) and Partial Autocorrelation Functions (PACFs) against the lag length, is used to identify the significant ACFs and PACFs. The lags of ACFs and PACFs, with probability values less than 5%, are significant and identified. Possible models are developed from these plots for the S&P and HSI index returns series. We estimate the ARIMA model for every in-sample period for both the S&P and HSI. The best model for forecasting is identified by considering information criteria, including Akaike Information Criteria (AIC) and Schwarz Bayesian Information Criteria (SBIC). It is also an accepted statistical paradigm that the correctly-specified model for the historical data will also be the optimal model for forecasting. Hence, it is reasonable to compare the best ANN results with those of ARIMA.

In this study, one of the most widely-used ANN models, called the feed-forward neural network, is used for financial time series forecasting. Usually, the ANN model consists of an input layer, an output layer and one or more hidden layers. The hidden layers can capture the nonlinear relationship between variables. Each layer consists of multiple neurons connected to neurons in adjacent layers.

The ANN Model

A pictorial representation of the ANN, as shown in Fig 3.1, illustrates the architecture of a neural network, with one hidden layer containing two neurons, three input variables and one output variable. This feed-forward or multi-perceptron network with one hidden layer is the most basic and commonly-used neural network in economic and financial applications.

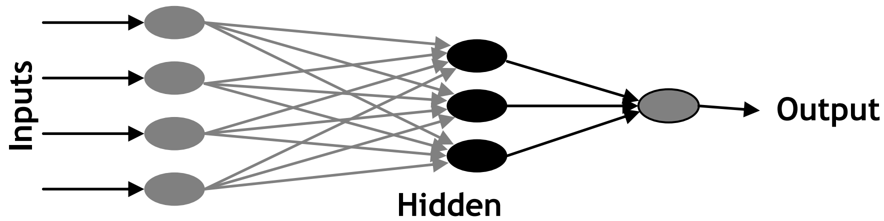

Figure 3.1 Single hidden layer feed-forward network

The first or lowest layer is called the input layer, which receives information or input signals. The last or highest layer is called the output layer, which is where the solution is obtained. The input and output layers are separated by one or more layers called the hidden layers. The neural network is said to be fully connected in the sense that every node in each layer of the network is connected to the every node in the adjacent forward layer. Every connection in a neural network has a weight associated with it. The ANN is trained in such a way to minimise the difference between the network output and the target. Therefore, training is the process of weight adjustment in order to obtain a desirable outcome. In other words, one wants the set of weights that produces the least squares residuals.

An ANN can be trained using the historical data of a time series to capture the nonlinear characteristics of that specific time series. The model parameters (i.e., connection weights and node biases) are adjusted iteratively by minimising the forecasting errors. For time series forecasting, the final computational form of the ANN model is:

$$Y_t = ao + \sum_{j=1}^{q} w_j f(a_j + \sum_{i=1}^{p} w_{ij} Y_{t-i}) + \varepsilon_t \tag{2}$$

where $a_j (j = 0, 1, 2, ..., q)$ is a bias on the j^{th} unit, and $w_{ij} (i = 1, 2, ..., p; j = 1, 2, ..., q)$ is the connection weight between layers of the model. $f(.)$ is the transfer function of the hidden layer. p is the number of input nodes, and q is the number of hidden nodes. Actually, the ANN model in (2) performs a nonlinear functional mapping from the past observation $(Y_{t-1}, Y_{t-2}, ..., Y_{t-p})$ to the future value Y_t. In other words,

$$Y_t = \varphi(Y_{t-1}, Y_{t-2}, Y_{t-3}......., Y_{t-p}, \nu) + \xi_t \tag{3}$$

where v is a vector of all parameters, and φ is a function determined by the network structure and connection weights. Thus, in some sense, the ANN model is equivalent to a nonlinear autoregressive (NAR) model.

This study employs a three-layer feed-forward ANN to forecast daily returns of the S&P 500 and HSI. The most important aspect of the model is the determination of the architecture of ANN for a particular application. The choice of input variables (i.e., nodes) and hidden nodes, activation functions, the learning rate, the number of epochs, the training algorithm and termination criteria are very crucial in determining the model's performance.

Theoretically, ANN models can approximate any complex function; in this context, the financial time series models can be fairly well approximated.

The Experimental Set-up

Manish and Thenmozhi (2007) used ANN to forecast India's S&P CNX Nifty Index. As presented below, an experimental set-up that follows their study is used to develop the ANN model.

Data

1. Data set: Daily returns on the S&P 500 and HSI Indices.
2. Selection of training and test data: Explained earlier (Dynamic Increasing Methodologies for Estimation and Prediction).
3. Input Variables: Past one-, two- and three-day returns.
4. Output Variables: One-day ahead returns.

Network

5. Topology: The three time series use one hidden layer with the following number of nodes in each layer: 3 TS-(1-4) LS-1L.
   Activation Function: TS-Tan sigmoid, L-Linear, with the number of levels hidden nodes experimented varying from one to five. All nodes in the network are fully connected without shortcut connections.
6. Initialisation: Fixed randomly-generated initial weights between –1.0 and 1.0 with random distribution.
7. Learning rate: .01.

Training

8. Training algorithm: Scaled Conjugate Gradient (SCG).
9. Termination criteria: 2,500 epochs that are fixed and linked to the data set, or the mean square error of MSE=0.
10. No. of runs of each simulation: Fixed at 15.

Analysis

11. Performance metrics: Fixed HR, annualised returns, Sharpe ratio, annualised volatility and maximum drawdown.

The study uses MATLAB 6.5 to build and train the ANN. The MATLAB program works with default parameter values of weight.

Out-of-Sample Performance Measures

HR

The ability of the models to predict the direction of the indices is evaluated by testing their market timing ability. This ability is measured by HR as calculated as the proportion of times the sign of the stock index is predicted correctly over the whole forecast period. The competing models will have market timing ability if the value of HR is greater than 50%. Mathematically, the prediction performance 'P' as measured by HR is evaluated using:

$$P = \frac{1}{m}\sum_{i=1}^{m} R_i$$

where R_i is the prediction result for the i^{th} trading day and is defined as:

$$R_i = \begin{cases} 1 \text{ if } PO_i = AO_i \\ 0 \text{ otherwise} \end{cases}$$

where PO_i is the predicted output from the model for the i^{th} trading day; AO_i is the actual output for the i^{th} trading day; and m is the number of test examples.

Trading Measure-Profit and Loss

Presented here are the operational details of the trading strategy. The seed money, which is $100 in this experiment, is used to buy stock index funds when the prediction shows a rise in the index price. To calculate the profit, index funds are bought or sold at the same time. It should be noted that the price of the index fund is directly proportional to the index level; a virtual investor can gain from both the rise and fall of the index price. The trading strategy is to go long when the model predicts that the index price will rise, that is, the forecast is positive; otherwise, the investor should sell. If the signal points to sell or buy for two or more consecutive days, the zero position is taken from the second day onwards. This situation means index funds are held until the next turning point that the model predicts. Thus, the positions are assumed to be a) zero, b) long for 100% of the fund value at the time of entry or c) short for 100% of the fund value at the time of entry. Exit positions are assumed to be full exits. The trading performance measures used to analyse the forecasting techniques include the monthly P&L returns, Sharpe ratio, annualised volatility and maximum drawdown.

RESULTS

ARIMA Model

To develop the ARIMA model, it is essential to check the stationarity of the time series. This study employs the Augmented Dickey and Fuller (ADF) test and the Philip Perrons (PP) test to check for stationarity in the HSI and S&P 500 returns time series. The results of these tests suggest that the log first difference for the two series is stationary.

The correlogram, which simply plots ACFs and PACFs against the lag length, was used to identify the significant ACFs and PACFs. Information criteria (i.e., AIC and SBIC) was used to identify the best forecasting model. After considering all possible models and looking at the AIC and SBIC values of each model, ARIMA (2,1,2) and ARIMA (1,1,1) were identified as the best models for forecasting daily returns on the S&P and HSI returns, respectively. Sometimes, the two criteria (i.e., AIC and SBIC) contradict. Empirically, a researcher usually prefers to use SBIC criteria for model selection. However, in this present study, the results of both AIC and SBIC criteria clearly identified the same ARIMA model for the two series (results available upon request).

To test for serial correlation in the residuals, we used the Breusch-Godfrey LM Test. The LM test for serial correlation of residuals suggests that the ARIMA models for the two series capture the entire serial correlation; the residuals do not exhibit any serial correlation. Thus, it is not necessary to search for another ARIMA model (Gujarati, 1995) (results are available upon request).

Neural Network Model

The architecture that achieves the best results in terms of generalisation had three inputs (i.e., three past observations of the S&P 500), four hidden neurons and one output to forecast the daily returns on the S&P 500 series. However, for the HSI returns, the architecture used three inputs (i.e., three past observations of HSI returns), three hidden neurons and an output to forecast the daily returns.

Forecast Evaluation

The experiment was done for the daily S&P 500 and HSI returns series, and a one-period ahead forecast was produced for the 36 test samples using the ANN and ARIMA models. The study measured the performance metrics on the test data to investigate how well the models captured the underlying trends of the movement of the time series. Tables 1 and 2 outline the predictive performance of the two models for the 36 test samples and for the two returns series.

Table 1
Prediction accuracy of the S&P 500.

Hit Ratio Using ANN and ARIMA Models for S&P 500 Returns						
Years	2006		2007		2008	
Months	ANN	ARIMA	ANN	ARIMA	ANN	ARIMA
January	60.00%	45.00%	55.00%	55.00%	38.10%	28.57%
February	63.16%	36.84%	47.37%	52.63%	70.00%	60.00%
March	34.78%	47.83%	63.64%	36.36%	35.00%	45.00%
April	68.42%	57.89%	65.00%	75.00%	59.09%	45.45%
May	45.45%	50.00%	63.64%	45.45%	52.38%	57.14%
June	59.09%	54.55%	47.62%	52.38%	42.86%	47.62%
July	50.00%	50.00%	52.38%	57.14%	63.64%	54.55%
August	43.48%	47.83%	56.52%	43.48%	71.43%	76.19%
September	55.00%	70.00%	63.16%	36.84%	71.43%	52.38%
October	81.82%	54.55%	52.17%	21.74%	34.78%	47.83%
November	66.67%	61.90%	47.62%	23.81%	47.37%	57.89%
December	55.00%	70.00%	65.00%	45.00%	59.09%	18.18%

For the S&P 500 returns series and the ANN Model, Table 1 suggests that in 2006, HR was over 51% in 8 out of 12 out-of-sample periods, while it was 9 in 2007 but 7 in 2008. The HR was over 55% in 6 out of 12 out-of-sample periods for 2006, 2007 and 2008. Moreover, the HR was over 60% in 4 out of 12 out-of-sample periods in 2006, 5 out of 12 out-of-sample periods in 2007 and 4 out of 12 out-of-samples in 2008.

For the ARIMA model, the results suggest that HR was over 51% in 6 months of 2006, 5 in 2007 and 6 in 2008. The HR was over 55% in 4 months in 2006, 2 in 2007 and 4 in 2008. Moreover, the HR was over 60% in only 3 months in 2006, 1 in 2007 and 1 in 2008.

Comparing the results of the HR metrics on the S&P 500 returns for ANN with that for ARIMA, it can be concluded that ANN outperformed ARIMA in 2006, 2007 and 2008.

Table 2
Prediction Accuracy of the HSI

Hit Ratio Using ANN and ARIMA Models for HSI Returns						
Years	2006		2007		2008	
Months	ANN	ARIMA	ANN	ARIMA	ANN	ARIMA
January	68.42%	57.89%	59.09%	54.55%	63.64%	50.00%
February	70.00%	65.00%	55.56%	55.56%	36.84%	68.42%
March	47.83%	43.48%	45.45%	50.00%	47.37%	42.11%
April	64.71%	58.82%	50.00%	38.89%	47.62%	47.62%
May	45.00%	50.00%	57.14%	47.62%	50.00%	35.00%
June	50.00%	50.00%	55.00%	50.00%	55.00%	65.00%

(continued)

Table 2 (*continued*)

Hit Ratio Using ANN and ARIMA Models for HSI Returns						
Years	2006		2007		2008	
Months	ANN	ARIMA	ANN	ARIMA	ANN	ARIMA
July	66.67%	52.38%	76.19%	52.38%	36.36%	50.00%
August	56.52%	52.17%	52.17%	65.22%	52.63%	31.58%
September	52.38%	47.62%	63.16%	52.63%	52.38%	38.10%
October	70.00%	70.00%	57.14%	52.38%	71.43%	57.14%
November	68.18%	59.09%	63.64%	59.09%	65.00%	40.00%
December	57.89%	47.37%	68.42%	63.16%	52.38%	52.38%

The HR results for the HSI returns in Table 2 suggest that in 2006, HR of ANN was over 51% in 9 out-of-sample periods, while it was 10 in 2007 and 7 in 2008. The HR was over 55 % in 8 out of 12 out-of-sample periods in 2006 and 2007 but decreased to 3 in 2008. Moreover, the HR was over 60% in 6 out of 12 out-of-sample period in 2006, while it was 4 in 2007 but decreased to 3 in 2008. The HR results for ARIMA model as shown in Table 2 suggest that in 2006, HR was over 50% in 7 out of 12 out-of-sample period, while it was 8 in 2007 but 4 in 2008. Thus, for the HSI returns as well, ANN outperformed ARIMA for the three years under analysis.

Results in Tables 1 and 2 are as expected from a linear model such as ARIMA. It is widely known that stock market data often contain noisy information and have nonlinear and chaotic dynamics. Thus, the S&P 500 and HSI returns as predicted by the ARIMA model do not provide a reasonable description of the asset price movement.

Trading Performance Results

The performance of the ANN model is encouraging. However, predictability does not necessarily imply profitability. Rather, a model that can assure profitability through the use of a particular type of strategy must be identified. To evaluate the performance of the two models, a simple strategy discussed in an earlier section (Trading Measure-Profit and Loss) was used in a trading simulation. Tables 3 and 4 summarise the results of the trading performance of the two models for the two time series.

The results of the ANN model for the S&P 500 series show that in 2006, the monthly P&L was positive for 9 out of 12 months and 7 in both 2007 and 2008. However, for the ARIMA model, the results of P&L for S&P 500 returns seem to be highly correlated with HR results. In 2006, the monthly P&L was positive for 6 of 12 months, whereas it was positive for 3 months in 2007 and 6 in 2008. In terms of the number of months the P&L is positive for the S&P 500 series, so it can be concluded that ANN performs better than ARIMA.

Table 3
Trading performance of S&P 500.

P&L Using ANN and ARIMA Models for S&P 500 Returns						
Years	2006		2007		2008	
Months	ANN	ARIMA	ANN	ARIMA	ANN	ARIMA
January	0.70%	–2.71%	1.38%	0.26%	–10.75%	–14.41%
February	1.29%	–0.58%	–2.26%	–2.89%	9.88%	4.90%
March	–0.43%	0.03%	4.05%	–1.35%	–5.22%	–0.96%
April	3.10%	0.53%	4.30%	3.84%	3.90%	–0.47%
May	–1.35%	–0.19%	3.22%	–2.23%	–0.91%	–2.33%
June	5.72%	–0.98%	–1.02%	–2.29%	–7.66%	–4.75%
July	–0.73%	–0.32%	–3.91%	–2.03%	6.36%	8.18%
August	0.56%	0.84%	–0.29%	1.25%	6.28%	7.82%
September	1.48%	1.63%	4.89%	–6.34%	32.16%	4.87%
October	4.28%	–0.11%	1.89%	–8.24%	2.72%	16.68%
November	2.50%	1.31%	–2.20%	–13.93%	–0.89%	25.82%
December	1.30%	4.58%	2.16%	–1.76%	19.61%	–27.95%

Table 4
Trading performance of HSI.

P&L Using ANN and ARIMA Models for HSI Returns						
Years	2006		2007		2008	
Months	ANN	ARIMA	ANN	ARIMA	ANN	ARIMA
January	5.83%	5.34%	2.67%	5.10%	39.39%	–3.54%
February	2.90%	1.41%	–1.63%	1.29%	–10.64%	15.25%
March	–2.02%	–2.52%	0.70%	–2.70%	–6.90%	–12.61%
April	5.34%	1.18%	2.37%	–0.45%	–0.13%	2.22%
May	–0.59%	-4.12%	5.35%	3.10%	1.54%	–11.29%
June	2.75%	13.88%	4.73%	3.42%	–3.25%	–2.81%
July	4.25%	1.45%	7.69%	1.12%	–10.13%	–1.44%
August	2.94%	3.21%	2.94%	19.92%	1.97%	–13.55%
September	0.67%	–1.00%	13.72%	7.09%	–0.11%	–4.69%
October	4.29%	3.73%	13.59%	5.75%	52.97%	0.17%
November	3.37%	0.54%	14.97%	15.66%	11.34%	–12.80%
December	7.16%	2.64%	14.31%	7.26%	0.70%	0.24%

The monthly P&L of the HSI returns for the ANN model is positive for 10 out of 12 months in 2006; it was positive for 11 months in 2007 but 6 in 2008. The results of P&L for the HSI returns using ARIMA seem to be highly correlated with HR results. In 2006, the monthly P&L was positive for 9 out of 12 months, while it was positive for 10 months in 2007 but 4 in 2008. In terms of numbers of months, the P&L is positive. So it can be concluded that the P&L of ANN is better than that of ARIMA in 2006, 2007 and 2008.

The results of the two models are also evaluated with other trading measures such as the Sharpe ratio, annualised volatility and maximum drawdown. Tables 5 and 6 summarise the results of the other trading performance measures of the two models and for the two time series on an annual basis.

Table 5
Other trading measures for S&P.

Trading Performance Using ANN and ARIMA Models for S&P Returns						
Years	2006		2007		2008	
Criteria	ANN	ARIMA	ANN	ARIMA	ANN	ARIMA
Sharpe ratio	2.4	1.88	3.8	1.25	2.55	1.9
Annualised volatility	9.9 %	12.0 %	15.6 %	16.2%	41.1 %	47.8 %
Maximum drawdown	–5.2 %	–7.13 %	–7.9 %	9.2 %	–32.1 %	–40.53 %
Index max drawdown	–7.7 %		–10.1 %		–48. 8 %	

Table 6
Other trading measures for HSI.

P&L Using ANN and ARIMA Models for HSI Returns						
Years	2006		2007		2008	
Criteria	ANN	ARIMA	ANN	ARIMA	ANN	ARIMA
Sharpe ratio	2.96	2.08	4.87	3.15	6.85	3.75
Annualised volatility	14.2 %	15.3 %	25.3 %	27.6 %	48.7 %	52.8 %
Maximum drawdown	–8.5 %	–10.3 %	–12.08 %	–16.1 %	–32.1 %	–48.05 %
Index max drawdown	–12.0 %		–17.8 %		–60.48 %	

The results of the ANN model for the two time series returns and for different out-of-sample periods in terms of risk-adjusted performance (Sharpe ratio), annualised volatility and other characteristics are quite impressive. It outperforms the ARIMA model. Moreover, the ANN model has the lowest downside risk as measured by maximum drawdown for the two series and for different sampling periods. As with statistical performance measures, financial criteria clearly single out the ANN model as the one with the most consistent performance; it is therefore considered the *best* model for this particular application. By considering P&L, financial criteria and HR, it can be concluded that it is advantageous to apply ANN to forecast financial time series.

Heteroscedasticity and Neural Network Model Performance

We have also investigated the forecasting performance of the two models after filtering out the volatility effects. Initially, we tested the two series for the ARCH effects. The results (available upon request) of the ARCH tests suggest that ARCH terms are present in both series. These results suggests that the performance of the models after removing the ARCH effects need to be reexamined. We performed the forecasting exercise using volatility-filtered series of S&P 500 and HSI returns. Tables 7 and 8 show the results.

Table 7
Volatility-filtered S&P 500 series.

		ANN	ARIMA
Annualised Hit Ratio	2006	54.18%	50.02%
	2007	56.45%	45.42%
	2008	52.72%	50.89%
Number of Months with Positive P&L	2006	9	5
	2007	5	3
	2008	5	4
Annualised Returns	2006	17.43%	3.82%
	2007	11.36%	–31.02%
	2008	45.51%	6.24%

Table 8
Volatility filtered HSI series.

		ANN	ARIMA
Annualised Hit Ratio	2006	57.89%	53.12%
	2007	57.72%	51.66%
	2008	52.15%	47.16%
Number of Months with Positive P&L	2006	9	8
	2007	10	8
	2008	4	3
Annualised Returns	2006	40.49%	19.70%
	2007	98.67%	76.00%
	2008	56.18%	–40.12%

Again, ANN offers reliable and consistent results both in terms of P&L and HR, regardless of market conditions. The results of ANN were much better than those of ARIMA for the two series.

With respect to forecasting stability, ANN is robust across all validation tests, and forecasting seems to be more stable. Thus, this study concludes that an ANN model predicts better in terms of HR and P&L.

CONCLUSION

Since the accuracy of forecasting is vital to many decision-making processes, improvements on the effectiveness of forecasting models have been perennially pursued. In this study, the relevance of ARIMA and ANN for forecasting the two time series has been examined by identifying and developing ARIMA- and

ANN-based models. The out-of-sample performance of these models is evaluated using HR, P&L and risk-adjusted measure such Sharpe ratio, annualised volatility and maximum drawdown for different out-of-sample periods. The results suggest that ARIMA does not provide better forecasts for the HSI and SPX time series, based on financial criteria, P&L and the HR ratio. Moreover, after removing the ARCH effect and reexamining the forecasting ability of the two models for the two series, the results show that ANN outperformed ARIMA not only in terms of HR but also for financial criteria. These results show that useful prediction is possible for the two series without using extensive market data or knowledge.

We suggest that traders develop models using ANN to forecast the different financial time series and to make better investment decisions. Financial forecasters, dealers and traders can benefit from different trading approaches based on ANN techniques. Future research may aim to integrate ANN with other models, whereby the weaknesses of one method can be balanced by the strengths of another. Future research may also assess the model's accuracy in conjunction with macroeconomic inputs, such as interest rates, consumer price index and industrial production, as well as technical indicators.

ACKNOWLEDGEMENTS

The author would like to thank the anonymous referee for his/her constructive suggestions that help to improve the manuscript.

REFERENCES

Ahmadi, H. (1990). Testability of the arbitrage pricing theory by neural networks. *Proceedings of the International Conference on Neural Network*, 385–393, San Diego, CA.

Azoff, E. M. (1994). *Neural network time series forecasting of financial markets*. New York: John Wiley & Sons.

Choi, J. H., Lee, M. K., & Lee, M. W. (1995). Trading S&P 500 Stock index futures using a neural network. Proceedings of the 3rd Annual International conference on artificial intelligence applications on Wall Street, 63–72, New York.

Corradi, V., & Swanson, N. R. (2005). Nonparametric bootstrap procedure for predictive inference based on recursive estimation schemes. Working Paper, Rutgers University.

Corradi, V., & Swanson, N. R. (2006). Bootstrap conditional distribution tests in the presence of dynamic misspecification. *Journal of Econometrics*, *133*(2), 779–806.

Donaldon, R. G., & Kamstra, M. (1996). A new dividend forecasting procedure rejects bubbles in asset prices: The case of 1929 stock crash. *Review of Financial Studies, 9*(2), 333–383.

Granger, C. W. J., & Newbold, P. (1986). *Forecasting economic time series*. San Diego: Academic Press.

Gujarati, D. (1995). *Basic Econometrics*. New York: McGraw-Hill.

Hornik, K., Stinchcombe, M., & White, H. (1989). Multilayer feedforward networks are universal approximators. *Neural Networks, 2*(5), 359–366.

Huang, W., Nakamori, Y., & Wang, Shou-Yang (2005). Forecasting stock market movement direction with support vector machine. *Computers & Operations Research*, *32*, 2513–2522.

Hwarng, H. B., & Ang, H. T. (2002). A simple neural network for ARMA (p,q) time series. *OMEGA: International Journal of Management Science*, *29*(4), 319–333.

Kim, K. (2003). Financial time series forecasting using support vector machines, *Neurocomputing*, *55*(1&2), 307–319.

Kim, K., & Han, I. (2000). Genetic algorithms approach to feature discretization in artificial neural networks for the prediction of stock price index. *Expert System Application*, *19*(2), 125–132.

Kimoto, T., Asakawa, K., & Takeoka, M. (1990). Stock market prediction system with modular neural networks. *Proceedings of the IEEE International Joint Conference on Neural Networks*, *2*, *11–16*. *San Diego, California,*

Leung, M. T., Daouk, H., & Chen, A. S. (2000). Forecasting stock indices: A comparison of classification and level estimation models. *International Journal of Forecasting*, *16*(3),173–190.

Manish, K., & Thenmozhi. M. (2007). An experimental analysis of selected training algorithms for neural networks in financial time series prediction. *International Journal of Operations and Quantitative Management, 13*(1), 1–18.

Refenes, A. N., Zapranis, A. S., & Francis, G. (1994). Stock performance modeling using neural networks: Comparative study with regressive models. *Neural Networks, 7*(2), 375–388.

Richard J. & Dean, W. (1998). *Applied Multivariate Statistical Analysis*. New Jersey: Prentice Hall.

Stock, J. H. & Watson, M. W. (2003). Forecasting output and inflation the role of asset prices. *Journal of Economic Literature*, *41*(3), 788–829.

Thenmozhi, M. (2001). Predictability of BSE returns using neural networks: An exploratory study. *Proceedings of International Conference on Industrial Mathematics, IIT Madras. India*.

Trippi, R. R., & D., DeSieno. (1992). Trading equity index futures with a neural network. *Journal of Portfolio Management*, *19*(1), 27–33.

Yaser, S. A. M., & Atiya, A. F. (1996). Introduction to Financial Forecasting. *Applied Intelligence*, *6*(3), 205–213.

Yu, L., Wang, S. Y., & Lai, K. K. (2005). A novel nonlinear ensemble forecasting model incorporating GLAR and ANN for foreign exchange rates. *Computers and Operations Research*, *32*(10), 2523–2541.

Zhang. (2003). Time series forecasting using a hybrid ARIMA and neural network model. *Neurocomputing*, *50*, 159–175.

Zhang, G., Patuwo, B. E., & Hu, M. Y. (1998). Forecasting with artificial neural networks: The state of the art. *International Journal of Forecasting*, *14*(1), 35–62.

Zirilli, J. (1997). *Financial prediction using neural network*. London: International Thompson Computer Press.