APPLICATION OF NEURAL NETWORKS TO AN EMERGING FINANCIAL MARKET: FORECASTING AND TRADING THE TAIWAN STOCK INDEX

Mark T. Leung

Division of Management and Marketing
College of Business
University of Texas
San Antonio, Texas 78249
U.S.A.

An-Sing Chen

National Chung Cheng University
Department of Finance
Ming-Hsiung, Chia-Yi
Taiwan, R.O.C 621

Hazem Daouk*

University of Michigan
Business School
701 Tappan Street
Room D3263
U.S.A.
Tel: 734-763-6391
Fax: 734-764-3146
E-mail: hdaouk@umich.edu
Website: webuser.bus.umich.edu/hdaouk

Forthcoming: Computers & Operations Research

First draft: January 9, 1998
Current draft: July 10, 2001

*Corresponding author

APPLICATION OF NEURAL NETWORKS TO AN EMERGING FINANCIAL MARKET: FORECASTING AND TRADING THE TAIWAN STOCK INDEX

ABSTRACT

In the last decade, neural networks have drawn noticeable attention from many computer and operations researchers. While some previous studies have found encouraging results with using this artificial intelligence technique to predict the movements of established financial markets, it is interesting to verify the persistence of this performance in the emerging markets. These rapid growing financial markets are usually characterized by high volatility, relatively smaller capitalization, and less price efficiency, features which may hinder the effectiveness of those forecasting models developed for established markets. In this study, we attempt to model and predict the direction of return on the Taiwan Stock Exchange Index, one of the fastest growing financial exchanges in developing Asian countries. Our approach is based on the notion that trading strategies guided by forecasts of the direction of price movement may be more effective and lead to higher profits. The Probabilistic Neural Network (PNN) is used to forecast the direction of index return after it is trained by historical data. The forecasts are applied to various index trading strategies, of which the performances are compared with those generated by the buy and hold strategy, and the investment strategies guided by the forecasts estimated by the random walk model and the parametric Generalized Methods of Moments (GMM) with Kalman filter. Empirical results show that the PNN-based investment strategies obtain higher returns than other investment strategies examined in this study. The influences of the length of investment horizon and the commission rate are also considered.

Keywords: Emerging Economy, Forecasting, Trading Strategy, Neural Networks, Generalized Methods of Moments (GMM).

1. Introduction

Although there exists some studies which deal with the issues of forecasting stock market index and development of trading strategies, most of the empirical findings are associated with the developed financial markets (e.g., U.S., U.K., and Japan). Currently, many international investment bankers and brokerage firms have major stakes in overseas markets. Given the economic success of Taiwan in the last two decades, the financial markets in this Asian country have attracted considerable global investments. This view is further corroborated by the recent introduction of several Taiwan Stock Index instruments by the Singapore International Monetary Exchange (SIMEX) in January 1997. Realizing the growing importance of the Taiwanese stock market and its influence on the current Asian financial crisis, our study attempts to develop effective forecasting models for predicting the Taiwan Stock Index returns. There are two basic reasons for a closer examination of this index trading vehicle. First, it provides an effective means for the investors to hedge against potential market risks. Second, it creates new profit making opportunities for market speculators and arbitrageurs. Therefore, being able to accurately forecast stock market indices has profound implications and significance to researchers and practitioners alike.

Another motivation for this study is to confirm whether we can extend some basic notions of traditional financial forecasting modeling, which are built upon the observation of well established financial systems, to a rapid growing emerging economy. Champion (1998) presents two diametrically opposed views of the Taiwanese market which differentiate it from the more developed financial markets. The author argues that "this market was different and had an internal logic of its own which allowed it to defy laws applicable elsewhere." However, there is an opposing view such that "normal financial relationships must sooner or later prevail." Therefore, given the rising popularity of index trading, it is of practical interest to assess the predictive strength of those explanatory variables, which are found to be useful in the forecasting of well-established markets, in Taiwan stock market.

Our study models and predicts the TSE Index using neural networks. Their performance is compared with that of parametric forecasting approaches, namely the Generalized Methods of

Moments (GMM) and random walk. The reasons for choosing those two models for comparison are as follows. The GMM is adopted because it is widely used in modeling financial time series. It is flexible, free of distributional assumptions, and was shown to be useful in forecasting. Random walk is also used because it is a natural benchmark that is based of the efficient market hypothesis. Many sophisticated forecasting models are not able to outperform the "naive" random walk model. To provide a more complete evaluation of the models, our comparison is based on not only the performance statistics but also the trading profits. Thus, this study develops a set of trading strategies to translate the forecasts into monetary returns. In addition, the experimental analysis investigates whether the length of the investment horizon has a significant impact on the quality of the forecasts. The remaining portion of this paper is organized as follows: A literature review and economic justification are given in the next section. In Section 3, we provide a description and conceptual foundation of the forecasting approaches (models) used in this study. Then, the results of forecasting are presented and discussed in Section 4. Section 5 describes the proposed index trading strategies which are driven by the forecasts made by various forecasting models. The last section concludes the paper.

2. Background

2.1 Evidence of Return Predictability

There exists considerable evidence showing that stock returns are to some extent predictable. Most of the research is conducted using data from well-established stock markets such as the U.S., Western Europe, and Japan. It is, thus, of interest to study the extent of stock market predictability using data from less well-established stock markets such as that of Taiwan.

For the U.S., several studies examine the cross-sectional relationship between stock returns and fundamental variables. Variables such as earnings yield, cash flow yield, book-to-market ratio, and size are shown to have some power predicting stock returns. Banz and Breen (1986), Jaffee, Keim and Westerfield (1989), and Fama and French (1992) are good examples of this group of research. Further, studies based on European markets report similar findings. The results of Ferson and Harvey (1993) indicate that returns are, to a certain extent, predictable across a number of European markets (e.g., U.K., France, Germany). In their study which is aimed at forecasting the U.K. stock prices, Jung and Boyd (1996) report “reasonably good” performance of their forecasts, suggesting that the predictive strength of their stock return models are not negligible. For the Japanese stock market, the empirical investigations by Jaffe and Westerfield (1985) and Kato, Ziemba, and Schwartz (1990) also find some evidence of predictability in the behavior of index returns.

Using time series analysis, Fama and French (1993) identify three common risk factors, an overall market factor, and some factors related to firm size and book-to-market equity which seem to explain the average returns on stocks and bonds. Moreover, Fama and Schwert (1977), Rozeff (1984), Keim and Stambaugh (1986), Campbell (1987), Fama and Bliss (1987), and Fama and French (1988a,b, 1990) find that macroeconomic variables such as short-term interest rates, expected inflation, dividend yields, yield spreads between long and short-term government bonds, yield spreads between low grade and high grade bonds, lagged price-earnings ratios, and lagged returns have some power to predict stock returns. At the same time, the studies by Chen, Roll, and Ross (1986) and Chan, Chen, and Hsieh (1985) suggest that changes in aggregate production, inflation, the short-term interest rates, the slope of term structure (measured by the difference in returns on long-term and short-term government bonds) and the risk premium (measured by the difference in returns on low grade and high grade bonds) are other macroeconomic factors that have some power to predict stock returns.

Although most of the papers in this avenue of research are related to the financial markets in developed economies, several recent articles do show that return predictability also exists in those less developed financial markets. Ferson and Harvey (1993) examine 18 international equity markets, some of which are found in developing economies. The study provides evidence of returns predictability. Harvey (1995) focuses on emerging markets by looking at the returns of more than 800 equities from 20 emerging markets including Taiwan. He finds that the degree of predictability in the emerging markets is greater than that found in the developed markets. In addition, local information plays a much more important role in predicting returns in the emerging markets than in the developed markets. This characteristic helps explaining the difference in predictability between the two kinds of markets.

2.2 Economic Rationale

In light of the previous literature, it is hypothesized that various measures of the macroeconomic environment which is available to the forecaster may be used as input state variables in the construction of prediction models to forecast the direction of movement of the stock market index. Table 1 outlines an array of such macroeconomic state variables which are applied to the paper. In the following, we will describe the economic intuition concerning why the state variables chosen in this study are expected to indicate future stock market movement.

The term structure of interest rates (TS), i.e., the spreads of long-term bond yields over short-term bond yields, may have some power to forecast stock returns. The hypothesis that this variable may have some power in forecasting stock returns is supported by the observation that this variable has a business cycle pattern. It is low around business peaks and high around business troughs. Thus, the term structure of interest rates captures the cyclical variation in expected returns. This fact, combined with the historical evidence which shows that stock returns are generally lower during recessions, substantiates the notion that term spread may exhibit some degree of predictive power on stock returns. This is because a large term spread may suggest probable business expansion or increased economic activity in the future that corresponds to higher stock returns. In short, the term spread variable may be thought of as an indicator of the future level of economic activity which then, indirectly, results in some power to forecast stock returns.

Short term interest rates also fluctuate with economic conditions. T-bill rates (TB) tend to be low in a business contraction, especially at the low turning points of business cycles. Therefore, low T-bill rates may indicate the future business expansion or increased economic activity to certain extent. Business expansions or increased economic activity has been historically associated with higher stock returns and recessions with lower stock returns. Like the term structure variable, the short term interest rate may also be thought of as an indicator of the future level of economic activity which then, indirectly, results in some power to forecast stock returns.

The lagged index return is included in this study to check whether the time series properties of the past index returns contain any information that is useful in forecasting the future index returns. The variables GC (Government Consumption), PC (Private Consumption), GNP (Gross National Product), GDP (Gross Domestic Product), CPI (Consumer Price Index) and IP (Industrial Production) are also included in our examination as they possibly contain imperative information concerning the forecast of future stock index returns.

2.3 Forecasting the Direction of Index Return

Most trading practices adopted by financial analysts rely on accurate prediction of the price levels of financial instruments. However, some recent studies have suggested that trading strategies guided by forecasts on the direction of price change may be more effective and generate higher profits. Wu and Zhang (1997) investigate the predictability of the direction of change in the future spot exchange rate. In another study, Aggarwal and Demaskey (1997) provide evidence that the performance of cross-hedging improves significantly if the direction of changes in exchange rates can be predicted. Based on the S&P 500 futures, Maberly (1986) explores the relationship between the direction of interday and intraday price change. O'Connor, Remus, and Griggs (1997) conduct a laboratory-based experiment and conclude that individuals are showing different tendencies and behaviors for upward and downward series. Finally, in their study on the All Ordinaries Index futures traded at the Australian Associated Stock Exchanges, Hodgson and Nicholls (1991) suggest to hold an evaluation of the economic significance of the direction of price changes in future research. In summary, the findings in these studies are reasonable because an accurate point estimation, as judged by its deviation from the actual observation, may not be a good predictor of the direction of change in the instrument's price level. Also, predicting the direction is a practical issue which usually affects a financial trader's decision to buy or sell an instrument.

3. Predicting Returns on Taiwan Stock Index

3.1 Data

The data used in this study are obtained from the E.P.S. database maintained by the

Department of Education of Taiwan. The data set covers the horizon from January 1982 to August 1992 and is divided into two periods: the first period runs from January 1982 to August 1987 and the second period runs from September 1987 to August 1992. The first period, the in-sample estimation period, is used for model determination (i.e., specifying the model parameters) and validation. The second period is the reserved out-of-sample evaluation period and is used to compare the forecasts and trading performances of various models. Depending on the length of the investment horizon, the forecasted variable is the continuously compounded three month (MR3), six month (MR6), or twelve month (MR12) excess returns on the Taiwan index. The independent variables (see Table 1) for predicting the index returns are all observable on or before the last day of the month preceding the month corresponding to the first day of the forecast period. For instance, for the prediction of the six month ahead continuously compounded return starting from the first day of March 1984, all independent variables must be observable on or before the last day of February 1984. Constructing the data set in this manner ensures that the generation of out-of-sample forecasts will be similar to those made in the real world. This is because only observable, but not future unobservable, data are used as inputs to the forecasting models.

3.2 Neural Network Forecasting

Both academic researchers and practitioners have made tremendous efforts to predict the future movements of stock market and devise financial trading strategies to translate the forecasts into profits. Recently, in addition to econometric forecasting approaches, artificial neural networks (ANN) have been demonstrated to provide promising results in financial forecasting and trading. A comprehensive review of the fundamental concepts and principals of the ANN can be found in Rumelhart and McClelland (1986) and Caudill and Butler (1993). Morever, Hawley, Johnson, and Raina (1990) and Medsker, Turban and Trippi (1993) provide an overview of the neural network models in the fields of finance and investment.

As stated in the introduction, emerging markets data exhibit properties that differ from those of developed markets. It is therefore unsatisfactory that very little has been done to date, in applying ANN to forecast emerging financial markets. There are however two notable exceptions. The first one is McNelis (1996). The paper examines the reaction of Brazilian stock prices to "shocks" in the U.S. and in Latin American markets, with classical linear methods, as well as with ANN. However, no attempt is made to forecast the stock prices. The second one is Chattopadhyay (1997). In that paper, a new methodology for predicting country risk ratings in evaluating global portfolio investment decisions is proposed. ANN are used to predict country risk ratings. The two papers above find the ANN approach to be quite useful. In our paper, we want to explore the usefulness of the ANN approach in forecasting an emerging market stock index.

3.2.1 Probabilistic Neural Network

The neural network models used in this study is based on the topology of Probabilistic Neural Network (PNN) proposed by Specht (1988, 1990). To be specific, PNN is a classifier in the sense that the network is able to deduce the class/group of a given input vector after the training process is completed. Interested readers should refer to the Appendix and Wasserman (1993) for detailed description of the PNN.

A number of appealing properties of the PNN justify our adoption of this type of neural network in this study. First, the training of PNN is rapid, enabling us to develop a frequently updated training scheme. The network is re-trained each time the data set is updated and thus the most current information can be incorporated into the network. Second, the network is able to identify outliers and questionable data points and thereby the extra effort on scrutinizing the training data can be reduced. Third, and the most important, the PNN provides the Bayesian probability of the class that a particular input belongs to. This valuable information is utilized by our trading strategies to determine the possible action. The actual implementation will be discussed in details later.

3.2.2 PNN Logic

PNN is conceptually built on the Bayesian method of classification which, given enough data, is capable of classifying a sample with the maximum probability of success (Wasserman (1993)). The principle of a Bayesian classifier rests on the selection of class i with the largest product term in the Bayesian Classification Theorem:

$$\max_{i}\{h_i \, l_i \, f_i(X)\} \tag{1}$$

where h_i is the *a priori* probability for class i,

l_i is the loss incurred by misclassifying a sample which truly belongs to class i,

$\boldsymbol{X}$ is $(x_1, x_2, \ldots., x_k)$, the input vector to be classified, and

$f_i(\boldsymbol{X})$ is the probability of $\boldsymbol{X}$ given the density function of class i.

It should be pointed out that the loss due to misclassification, in most cases, cannot be observed or measured in the sample data. To avoid unnecessary complexity, it is assumed that the loss l_i is the same for all classes in our study. The PNN-guided trading strategies proposed in this study can explicitly take into account of the loss (opportunity cost of misclassification) associated with each class.

As seen in Equation 1, the Bayesian decision rule requires a knowledge of the probability density functions of the classes. These density functions are directly estimated from a set of training samples using Parzen's window approximation method (1962). The PNN used in this study applies the Cacoullos's (1966) multivariate extension of the original Parzen density estimation which allows us to generate the joint density functions for a set of k variables. Equation 2 shows the expression:

$$f_i(X) = \frac{1}{(2\pi)^{k/2}\sigma^k n_i} \sum_{j=1}^{n_i} e^{\frac{-(X-Y_{ij})'(X-Y_{ij})}{2\sigma^2}} \tag{2}$$

where $\boldsymbol{X}$ is the input vector to be classified,

k is the number of variables in the input vector $\boldsymbol{X}$,

n_i is the number of training samples which belongs to class i,

$\boldsymbol{Y}_{ij}$ is the j^{th} training sample in class i, and

s is a smoothing parameter.

Interested readers should refer to the Appendix for a complete description of the PNN logic and mathematical details.

The topology of PNN uses three kinds of processing units (pattern, class, and output processing units) to implement the classification logic. During supervised training, a pattern processing unit is placed in the PNN to represent the input-output pattern presented by each training sample. Also, corresponding connections linking the processing unit to the class it belongs are established. The network continues this "learning process" until all training samples have been examined. Therefore, our PNN model contains a total of 68 pattern processing units, one for each sample in the training set. In addition, there are two class processing units, representing the upward and downward directions of index returns. The values computed by these units are forwarded to the output processing units which determine the class affiliation. Figure 1 depicts the actual PNN architecture for making the first out-of-sample forecast (Period 69) with three month investment horizon (MR3). With four independent variables in the input vector (TB, GC12, GNP12, and GDP6), the output of this network should indicate not only the predicted direction (sign) of the index return but also the Bayesian probability of the class affiliation. As mentioned earlier, our trading strategies will use this probability to make the asset allocation decision.

3.2.3 Training and Testing Scheme

A rolling horizon approach similar to Refenes (1993) is applied to the training of our neural networks. This approach updates the training set for every out-of-sample prediction and thus, theoretically, incorporates the latest observed information into the network. After an out-of-sample forecast is made, the entire training set slides forward for one period and the same network training procedure is repeated. To make the first forecasts, the 68 in-sample observations (from Periods 1 to 68) are used for training. Then the trained network predicts the direction of the index return in Period 69. After the prediction is made and recorded, the training set then slides forward for one period (covering Periods 2 to 69) and the process of training and testing is carried out again. As a result, 60 out-of-sample forecasts on the direction of index return are generated for the purpose of testing. This training and testing scheme is applied to all three month (MR3), six month (MR6), and twelve month (MR12) models [1].

3.3 GMM-Kalman Filter Forecasting

Since Hansen (1982), GMM has been one of the most popular econometric models for financial time series. First, GMM is very flexible and nests many more traditional econometric models as special cases. Second, GMM does not require distributional assumptions. This is particularly useful when dealing with financial data, especially emerging markets data. There is strong evidence in the literature that returns from emerging markets have empirical distributions that are very difficult to fit with commonly used statistical distributions. Third, Harvey (1995) has used GMM in his study of the forecastability of emerging markets stock returns and has found it to be very useful. Fourth, more recently, Chen and Chiou (1998) have used GMM to forecast the returns of an emerging market and have found it to perform very well. Therefore, besides neural networks, Generalized Method of Moments (GMM) with Kalman filter is used to generate forecasts. There are 60 out-of-sample forecasts for each of the three month, six month, and twelve month investment horizons, providing an equal basis for comparison with the neural network models. The specifications for the GMM models are based on the input variables determined by the FPE minimization procedure described in Table 2. The Kalman filter estimation method is an updating method which bases the model estimates for each time period on last period's estimates plus the data for the current time period; that is, it bases its estimates only on data up to and including the current period. This makes it highly useful for constructing forecasts which are based only on historical data. When applied to a standard linear model or in our case the GMM model, it provides a convenient way to compute a new coefficient vector when an additional observation is revealed.

The Kalman filter used in this study can be written in the following way: Let β_t denote the vector of *states* (coefficients) corresponding to the state variables at time *t*. The measurement equation is the GMM model:

$$y_{t+1} = X_t \beta_t + \mu_t \tag{3}$$

where y_t is the dependent variable and there are m independent variables or columns in the matrix $\boldsymbol{X_t}$. The variance of μ_t is $\boldsymbol{n_t}$. The state vector follows the process

$$\beta_t = \beta_{t-1} + \nu_t \tag{4}$$

with . $Var(\nu_t) = M_t$. μ_t and ν_t are independent. n_t and $\boldsymbol{M_t}$ (variance of the change in state vector) are assumed to be known. The Kalman filter recursively updates the estimate of β_t (and its variance), using the new information in y_t and $\boldsymbol{X_t}$ for each observation. Once we have an estimate of β_{t-1} and its covariance matrix Σ_{t-1} , then the updated estimate given y_t and $\boldsymbol{X_t}$ is

$$S_t = \Sigma_{t-1} + M_t \tag{5}$$

$$\Sigma_t = S_t - S_t X_t' (X_t S_t X_t' + n_t)^{-1} X_t S_t \tag{6}$$

$$\beta_t = \beta_{t-1} + \Sigma_t X_t' n_t^{-1} (y_t - X_t \beta_{t-1}) \tag{7}$$

o use the previous updating equations we need to supply β_0, the initial state vector; Σ_0, the initial covariance matrix of the states; n_t, the variance of the measurement equation; $\boldsymbol{M_t}$, the variance of the change in the state vector. In our study, $\boldsymbol{M_t}$, is set to 0, n_t, Σ_0, and β_0 are estimated using GMM. Details of the GMM estimation are described in the Appendix.

3.4 Random Walk Model

According to the Efficient Market Hypothesis (EMH), asset prices will follow a random walk as news is instantaneously incorporated into prices. The random walk model is therefore a natural theoretical benchmark. Random walk has been used as a benchmark for forecasting ability by numerous studies over the years. The random walk model is a very simple model to use and is often termed the “naïve model” because it does not involve much technical skills to implement. However, it has been shown to outperform, in terms of forecasting, many sophisticated methods. Therefore, it is a norm in the financial forecasting area to use it as a benchmark. The argument is that any new model that involves the implementation of advanced techniques should at least outperform the random walk model. Otherwise, the random walk model will be preferred since it does not involve much effort. For those reasons, we compare the performance of the PNN and GMM models with that of the random walk model. The random walk model assumes that the best forecast is equal to the most recently observable observation. Thus, the best three month ahead, six month ahead, and twelve month ahead forecast using the random walk model would be:

$$MR3_{t+1} = MR3_t \tag{8}$$

$$MR6_{t+1} = MR6_t \tag{9}$$

$$MR12_{t+1} = MR12_t \tag{10}$$

where $MR3_t$, $MR6_t$, and $MR12_t$ are the signs (i.e., directions) of the continuously compounded three, six, and twelve month annualized excess returns of the Taiwan index previous to time t. In other words, the random walk forecasts are the sign of the most recently observable returns corresponding to our forecast horizon.

3.5 Model Determination

An important aspect in developing forecasting models involves specifying the input variables. This can be done either by using: 1. economic arguments and theoretical reasoning to identify the principal determinants of the outputs; or 2. statistical testing procedure to justify the inclusion of particular input variables. In Section 2, we have given the economic arguments of why certain macroeconomic state variables may serve as inputs to our prediction models. Nevertheless, these economic arguments do not tell us how many lags of the input state variables should be included in the forecasting models. Further, establishing "pilot" models with a large number of input variables is not realistic.

Therefore, we adopt a statistical procedure using data from the in-sample estimation period (from January 1982 to August 1987) to narrow down the list of potential input state variables (see Table 1) for the prediction models and to calibrate the model specifications. The employed procedure, which is based on Akaike's minimum final prediction error (FPE), is similar to the ones used by Hsiao (1979) and Kaylen (1988). The FPE criterion is derived by assuming a quadratic loss function for each equation and is computed as follow:

$$FPE = \frac{\sum_{t=1}^{T} (Z_{it} - \hat{Z}_{it})^2}{T} \times \frac{T+N}{T-N} \tag{11}$$

where Z_{it} is the i^{th} independent state variable, N is the total number of parameters in the equation, and T is the number of observations. The first term on the right-hand side of Equation 11 is a measure of modeling error while the second term, *(T+N) / (T-N)*, is an adjustment for degrees of freedom.

Assuming that the k independent state variables are considered as one group, the procedure is as follows: First, consider the dependent variable *MR(d)* where d = 3, 6, 12. Then regress *MR(d)* on each independent state variable, one at a time, with lags from 1 to m using GMM. The FPEs are computed by varying the variables from 1 to k with lags from 1 to m. All of the independent state variables are then searched to find the state variable, Z_i , and its associated lags, p, which lead to the minimum FPE. This procedure is repeated for each of the remaining independent variables with an additional variable or lag entering the specification only if it reduces the FPE. Table 2 tabulates the results of these estimations.

As shown in Table 2, the statistical tests suggest a model specification for MR3 which includes TB, GC12, GNP12, and GDP6 as the explanatory state variables. The MR6 model specification includes TB, GC12, GNP12, GDP6, and CPI12 as the explanatory input variables. Finally, the MR12 model specification has the input vector consisting of variables TB, GC12, GNP12, GDP12, CPI12, and IP12. These model specifications are reported in Table 3 [2].

TB is expected to be negatively related to stock returns. The reason is that low T-bill rates may indicate a future business expansion or increased economic activity. Business expansions or increased economic activity has been historically associated with higher stock returns and recessions with lower stock returns. Whether the other state variables are positively or negatively correlated with future stock index returns is uncertain. To be specific, if these variables turn out to be reasonable proxies for the current health of the economy, then they will be negatively correlated with future index returns. This is because the economy is known to follow a cyclical pattern where periods of expansions are followed by periods of recessions. On the other hand, if these variables turn out to be reasonable proxies for the future growth rates of the economy, they will be positively correlated with future index returns.

4. Results

A total of 60 out-of-sample forecasts are made for the reserved period from September 1987 to August 1992. Table 4 tabulates the predicted direction of monthly index return by the PNN, the GMM-Kalman filter, and the Random Walk models for each of the out-of-sample periods. Actual (observed) index returns are also included for comparison.

Table 5 provides the summary statistics of the out-of-sample forecasts made by the PNN, GMM-Kalman filter, and random walk models. Shown on the table are the numbers of times (NC) the forecast correctly predicted whether the return is going to be positive or negative. It can be seen that the PNN model outperforms the others in predicting the direction of all three month ahead, six month ahead, and twelve month ahead index returns.

Further, with any length of the investment horizon, the PNN model is able to correctly predict the directions of excess returns more than 50% of the time at the 5% level of statistical significance. This implies that there is less than a 5% chance that these forecasting results may be due to chance alone. On the other hand, the GMM-Kalman filter model does not perform quite as well as the PNN model and is able to correctly predict the directions more than 50% of the time at the 10% level of statistical significance for only the six month ahead returns.

5. Trading Strategies and Experiment

The evidence that some of the forecasting methods documented in this study are able to correctly predict the direction of index return more than 50% of the time suggests that it may be possible to construct a set of economically profitable investment strategies. Therefore, we formulate a set of trading rules guided by the directions predicted by PNN, GMM-Kalman filter, and random walk models. The empirical testing takes the form of a trading simulation which closely mimics the timely investment decisions faced by investors in the marketplace. This trading simulation also allows us to evaluate the relative economic profit of the proposed investment strategies.

Essentially, the trading simulation investigates the influence of three experimental factors: 1. length of the investment horizon; 2. commission rates; and 3. investment strategies. The length of investment horizon is the period of time in which the index returns are realized. This is practically the same as the horizon lengths associated with the predicted index return direction. Thus, three month, six month, and twelve month investment horizons are used to implement the forecasts made by the MR3, MR6, and MR12 models, respectively. Three commission rates, 0.03%, 3%, and 6%, are examined in the experiment. It should be pointed out that the 0.03% commission rate is approximately the average transaction cost faced by a typical individual investor in Taiwan. Lastly, the economic performances of the trading strategies makes use of the out-of-sample forecasts made by PNN, GMM-Kalman filter, and random walk, and a simple buy and hold strategy are included for comparison.

5.1 Trading Simulation

We now describe the operational details of the trading simulation. The simulation experiment assumes that, at the beginning of each monthly period, the investor makes an asset allocation decision of whether to shift his liquid assets into the riskfree bonds or into the stock index fund. Liquid assets are defined as money that is currently not invested in either the riskfree bonds or the stock index fund. It should be noted that the price of the stock index fund is directly proportional to the index level. Further, it is assumed that the money that has been invested in either riskfree bonds or the stock index fund becomes illiquid and will not become liquid until the end of the investor's chosen investment horizon. In other words, the invested money will become available after the selected investment horizon reaches its maturity. For example, suppose the investor has decided to use an investment horizon of six months. The money that he has invested into either riskfree bonds or the stock index fund in the last six months is considered to have been "locked up" in that security. Hence, the asset will not be available for another round of investment decision before the security matures.

The testing period runs from September 1987 to August 1992 for a total of 60 months of out-of-sample observations. The first twelve periods of the simulation test period, however, are reserved for initialization of the trading simulation. In the trading experiment, it is assumed that, during the initiation period, an investor will invest $1 at the beginning of each month in either risk free bonds or the stock index fund depending on his chosen investment strategy. To achieve a fair comparison of the strategies with different investment horizons, the initialization period for the three month and six month investment plans are delayed so that the first maturity of any investment plan coincides with those of the twelve month investment plans which occurs at the end of Period 80. This is equivalent to the end of the 12th period in the simulation test period. The span of the initiation period varies for the different strategies in order to account for different investment horizon lengths before the first maturity. In particular, the investor will invest $1 at the beginning of each month in either riskfree bonds or the index fund from Periods 10 through 12 if his investment strategies are based on three month investment horizon. Similarly, he will invest the monthly $1 at the beginning of each month in risk free bonds or the index fund from Periods 7 through 12 and from Periods 1 through 12 for the six month and twelve month investment plans, respectively. Transaction costs, brokerage cost, or commissions are assumed to be incurred every time a transaction takes place.

5.2 Trading Strategies

PNN-Guided Trading Strategies

Given the superior performance of the PNN forecasts, we propose two PNN-guided investment strategies which translate the predicted direction of index returns into asset allocation decision. As mentioned earlier, a distinctive characteristic of PNN is its ability to provide the Bayesian probability associated with a classification. Let P_i denote the Bayesian probability for affiliation with class *i*. In our proposed strategies, P_i is compared against some established threshold. Practically, if P_i is higher than the threshold level, the asset will be allocated to the security tied up to class i. This single threshold triggering can be extended to a triggering strategy involving multiple thresholds.

Single Threshold Triggering

There is a single threshold in this naive version of the PNN-guided trading scheme. Without any prior knowledge on the market trend or subjective preference over a certain type of security, the threshold level is set to be at 0.5. In the simulation study, the investor allocates the assets to the riskfree government bonds when the predicted return is on the negative direction whereas he puts the assets into the stock index fund when the predicted return is positive. Hence, the assets will be allocated to the stock index fund if the probability of the predicted return is on the up trend is more than 0.5000. On the contrary, the assets will be allocated to the bonds if the respective probability is less than 0.5000. For the case that the probability is exactly equal to 0.5000, the asset allocation decision will follow that in the previous period.

Multiple Threshold Triggering

The single threshold triggering strategy ignores the asymmetric outcomes of the stock and bond markets. To be specific, the loss to misclassifying the upward movement as a downward one is not the same as the loss to misclassifying the downward movement as an upward one. If the PNN predicts an upward direction but the actual market is on the down trend, the investor will suffer a loss in the depreciation of the stock index fund. However, if the PNN predicts a negative return in the stock market but the actual market return is positive, the investor's assets will still appreciate slightly due to the interest paid on the maturity of the riskfree bonds. To take these asymmetric payoffs into consideration, we set up a two-threshold triggering strategy. The assets will be allocated to the stock index fund if the probability of a positive predicted direction of return is at least 0.7000 while the assets will be put into the bonds if the probability is less than 0.5000. The investor will keep his asset allocation decision unchanged if the probability is between 0.5000 and 0.7000. Mathematical representations of the single and multiple threshold triggers are presented below. Equation 12 signifies the decision rule for single threshold triggering at the 0.50 level whereas Equation 13 represents the one for multiple threshold triggering:

$$D_t = \begin{cases} 1 & for\ P_2 < 0.5 \\ D_{t-1} & for\ P_2 = 0.5 \\ 2 & for\ P_2 > 0.5 \end{cases} \tag{12}$$

$$D_t = \begin{cases} 1 & for\ P_2 < 0.5 \\ D_{t-1} & for\ 0.5 \leq P_2 \leq 0.7 \\ 2 & for\ P_2 > 0.7 \end{cases} \tag{13}$$

where D_t is the modified result of PNN classification of the input sample for Period t, and

P_2 is the PNN computed Bayesian probability that the input sample for Period t belongs to Class 2 (upward change in index level).

This investment strategy is intuitive. For a conservative investor, it would be better to invest in the riskfree bond market if the network is not certain about its prediction. In other words, the conservative investor should only invest in the riskier stock index when the network has a great confidence (i.e., a significant degree of certainty) on its prediction. The range between 0.5000 and 0.7000 represents a "buffer" region in the asset allocation decision making. There are two reasons for establishing such a buffer region. First, it minimizes the number of unnecessary transactions. Second, it reduces the chance of misclassification due to uncertainty.

<u>*Buy and Hold Strategy*</u>

The investor invests his money in the stock index fund and holds the fund till the end of the simulation test horizon, that is, the end of Period 128.

<u>*GMM-Kalman-Guided Strategies*</u>

The investor will follow the directions of returns predicted by the random walk models and GMM-Kalman filter models. Similar to the learning network based investment strategies, the investment strategies using these econometric models allocate the assets to the stock index fund when there is a predicted up trend and allocate the assets to the bonds when there is a predicted down trend.

5.3 Results and Analysis

The net gain in assets, number of trades executed, and the rate of return over the out-of-sample forecast horizon are shown in Table 6. Since the initial investments for various investment horizons are different, the percentage rate of return is the proper measure that can be compared across the scenarios.

$$\text{Rate of Return} = \frac{\text{Net Gain in Assets}}{\text{Initial Investment}} \tag{14}$$

From both Table 4 and Figure 2, it can be seen that both the single threshold (ST) and multiple threshold (MT) PNN-guided trading rules consistently outperform the ones guided by GMM-Kalman filter (KF), random walk (RW) forecasts and the buy and hold (BH) strategy. Also, the PNN-guided investment strategies with multiple triggering thresholds are generally better than those ones with single triggering threshold although the difference is marginal in some scenarios. This illustrates that the degree of certainty for PNN classification could have an impact on the interpretation of the predictions, especially when the investment horizon is long. A relatively longer investment horizon does not allow a quick switch of the underlying security and thus the strategies using multiple threshold triggering could reduce potential loss when the PNN is not certain about its prediction.

The buy and hold strategy (BH) results in a net loss when the investment horizon is three or six months. This is because the corresponding investment strategies, unlike the ones adopting the PNN and GMM-Kalman filter forecasts, are too rigid in that it is unable to capture the profits in the stock market and shift the realized profits into the bonds to preserve the asset worth when the market is on the down trend.

The random walk oriented investment strategy (RW) makes the asset allocation decisions solely based on the most recent information. Thus, it can serve as a benchmark alternative in which no information prior to the previous investment period is being incorporated into the forecasts. As shown in Table 6, the performance of RW drops by at least 50% when the investment horizon stretches from three months to six or twelve months. The observed sharp deterioration illustrates that this naive method which fails to take into account of the possible time series pattern is incapable of providing a reliable outlook to the more distant future. A comparison with the returns generated by PNN and GMM-Kalman filter trading strategies suggests that the timely historical information could be useful in predicting the return of the market. This notion is in agreement with many studies outlined in previous sections.

It follows the intuition that the rate of return for a given investment strategy decreases when the commission rate rises. However, it would be interesting to examine the change in the rate of return for a given trading strategy when the commission becomes more significant. From Table 6, for a particular trading strategy, the relative decrease in return is sharper for an increase in commission rates occurred in the three month plan than that in the six month or twelve month plans, implying that higher commission has a greater influence on the relatively shorter investment horizon. Figure 2 postulates that this observation is generally true for every trading strategy. A possible explanation is that the high commission harshly hampers the growth of the three month investment plans in early periods and thus limits the asset amount for reinvestment in later periods. However, for the plans with longer investment horizons, the assets are allowed to grow without transaction cost deduction for much longer periods of time during the early stage. This could provide a larger amount of assets for reinvestment growth in later time. In general, the investor should adopt a shorter investment horizon if the commission is low and a longer horizon if the commission is relatively higher.

The results show that the performance of both the PNN forecasts and the PNN-guided strategies are unusually strong as compared to what is normally expected from well-established financial markets. The readers should be aware that Taiwan was a strong emerging economy and experienced remarkable growth during the testing period. In fact, the artificially inflated “bubble” economy created by bullish investors could have made many financial figures quite forecastable. After that, its financial sector encountered a sharp downturn and this is explained by the poor performance of the buy and hold strategy. In addition, the PNN’s ability to estimate the underlying density function and map out the response surface may help explain its performance in such a volatile market.

6. Conclusions

The good performance of the PNN suggests that the neural network models are useful in predicting the direction of index returns. Furthermore, PNN has demonstrated a stronger predictive power than both the GMM-Kalman filter and the random walk forecasting models. This superiority is partially attributed to PNN’s ability to identify outliers and erroneous data. Compared to the other two parametric techniques examined in this study, PNN does not require any assumption of the underlying probability density functions of the class populations. Each density function is estimated by Parzen’s window approximation method.

The trading experiment shows that the PNN-guided trading strategies obtain higher profits than the other investment strategies utilizing the market direction generated by the parametric forecasting methods. In addition, the PNN-guided trading with multiple triggering thresholds is generally better than the one with single triggering thresholds. The multiple threshold version is able to consider the degree of certainty of a particular PNN classification and thereby reduce potential loss in the market.

A possible extension to enhance the PNN investment decision making is to include a set of adaptive thresholds, which changes dynamically in accordance with some opportunity cost. This can be achieved by setting the threshold levels with respect to the current and predicted interest rates and the price of the interest rate instrument.

Endnotes

[1] The training of the ANN takes less than 10 seconds. This is a very reasonable execution time. On the other hand, what is more time consuming is the selection procedure of the "best model". This refers to estimating different models and calculation diagnostics to assess their fit to the data and their forecasting ability during the validation stage. However, this is the same for the other econometric models. The ANN model selection procedure is not more time consuming than the selection procedure for the other econometric models.

[2] A quick inspection of these model specifications shows that several of the included macroeconomic input variables are likely to be collinear. However, for this study, multicollinearity in the input variables does not pose problems. First, none of the macroeconomic input variables is perfectly collinear and thus the resulting model specifications can still be estimated. Second, even in situations where multicollinearity is very high (near multicollinearity), the OLS estimators still retain the property of BLUE. This is because near multicollinearity per se does not violate the assumptions of the classical linear regression model. This means that unbiased, consistent estimates will be obtained, and thereby, their standard errors will be correctly estimated. The only effect of near multicollinearity is to make it hard to get coefficient estimates with small standard errors. It is shown in the Appendix that OLS is just a special case of the GMM where the residuals are homoscedastic and do not overlap. Thus, the properties of OLS with respect to near multicollinearity are also present when estimation is done using GMM instead of OLS. Finally, if the sole purpose of regression analysis, whether using GMM or OLS, is for forecasting rather than hypothesis testing, then near multicollinearity is not a serious problem. Also, using a larger number of independent variables that exhibit some collinearity may at times, produce better forecasts than a similar model that drops some of the independent variables in an attempt to reduce multicollinearity.

REFERENCES

Aggarwal, R. and Demaskey, A., 1997, Using Derivatives in Major Currencies for Cross-Hedging Currency Risks in Asian Emerging Markets, *Journal of Futures Markets*, 17, 781-796.

Banz, R. and Breen, W., 1986, Sample-Dependent Results Using Accounting and Market Data: Some Evidence, *Journal of Finance*, 41, 779-793.

Blattberg, R. and Gonedes, N., 1974, A Comparison of the Stable and Student Distributions as Statistical Models of Stock Prices, *Journal of Business*, 47, 244-280.

Cacoullos, T., 1966, Estimation of a Multivariate Density, *Annals of the Institute of Statistical Mathematics*, 18, 179-189.

Campbell, J., 1987, Stock Returns and the Term Structure, *Journal of Financial Economics*, 18, 373-399.

Caudill, M. and Butler, C., 1983, *Understanding Neural Networks: Computer Explorations* (MIT Press Cambridge, Massachusetts).

Champion, S. R., 1998, *The Great Taiwan Bubble* (Pacific View Press, Berkeley, California).

Chan, K., Chen, N., and Hsieh, D., 1985, An Exploratory Investigation of the Firm Size Effect, *Journal of Financial Economics*, 14, 451-471.

Chattopadhyay, S., 1997, Neural Network Approach for Assessing Country Risk for Foreign Investment, *International Journal of Management*, 14, 159-167.

Chen, A., Chiou, N., 1998, Forecasting the Direction of Stock Index Returns, *Review of Securities and Futures Markets,*

Chen, N., Roll, R., and Ross, S., 1986, Economic Forces and the Stock Market. *Journal of Business*, *59*, 383-403.

Fama, E. and Bliss, R., 1987, The Information in Long-Maturity Forward Rates. *American Economics Review*, *77,* 680-692.

Fama, E. and French, K., 1988a, Permanent and Temporary Components of Stock Prices, *Journal of Political Economy*, *96*, 246-273.

Fama, E. and French, K., 1988b, Dividend Yields and Expected Stock Returns. *Journal of Financial Economics*, *22*, 3-25.

Fama, E. and French, K., 1990, Business Conditions and Expected Returns on Stocks and Bonds. *Journal of Financial Economics*, *25*, 23-49.

Fama, E. and French, K., 1992, The Cross-Section of Expected Stock Returns. *Journal of Finance*, *47*, 427-465.

Fama, E. and French, K., 1993, Common Risk Factors in the Returns on Stocks and Bonds. *Journal of Financial Economics*, *33*, 3-56.

Fama, E. and Schwert, W., 1977, Asset Returns and Inflation. *Journal of Financial Economics*, *5*, 115-146.

Ferson, W. E. and Harvey, C. R., 1993, The Risk and Predictability of International Equity Returns. *Review of Financial Studies*, 6, 527-566.

Hansen, L., 1982, Large Sample Properties of Generalized Method of Moments Estimators. *Econometrica*, *50*, 1029-1054.

Hansen, L., and Hodrick, R., 1980, Forward Exchange Rates as Optimal Predictors of Future spot Rates, *Journal of Political Economy, 88*, 829-853.

Hartman, E., Keeler, J., and Kowalski, J., 1990, Layered Neural Networks with Gaussian Hidden Units as Universal Approximations. *Neural Computation*, *2*, 210-215.

Harvey, C. R., 1995, Predictable Risk and Returns in Emerging Markets. *Review of Financial Studies*, 8, 773-816.

Hawley, D., Johnson, J., and Raina, D., 1990, Artificial Neural Systems: A New Tool for Financial Decision-Making. *Financial Analysts Journal*, November/December, 63-72.

Hodgson, A. and Nicholls, D., 1991, The Impact of Index Futures Markets on Australian Sharemarket Volatility, Journal of Business Finance and Accounting, 18, 267-280.

Hsiao, C., 1979, Autoregression Modeling of Canadian Money and Income Data, *Journal of the American Statistical Association*, 74, 553-560.

Hutchinson, J., Lo, A., and Poggio, T., 1994, A Nonparametric approach to Pricing and Hedging Derivative Securities via Learning Networks. *Journal of Finance*, *49*, 851-889.

Jaffe, J., Keim, D., and Westerfield, R., 1989, Earnings Yields, Market Values and Stock Returns. *Journal of Finance*, *44*, 135-148.

Jaffe, J., and Westerfield, R., 1985, Patterns in Japanese Common Stock Returns: Day of the Week and Turn of the Year Effects. *Journal of Financial and Quantitative Analysis*, *20*, 261-272.

Jorion, P., 1995, Predicting Volatility in the Foreign Exchange Market, *Journal of Finance, 2*, 507-528.

Jung, C. and Boyd, R., 1996, Forecasting UK Stock Prices. Applied Financial Economics, 6, 279-286.

Kato, K. Ziemba, W., and Schwartz, S., Day of the Week Effects in Japanese Stocks. *Japanese Capital Markets*, ed. E. Elton and M. Grubber. New York: Harper and Row, 1990.

Kaylen, M.S., 1988, Vector Autoregression Forecasting Models: Recent Developments Applied to the U.S. Hog Market, *American Journal of Agricultural Economics*, 1988, *70*, 701-712.

Keim, D. and Stambaugh, R., 1986, Predicting Returns in the Stock and Bond Markets. *Journal of Financial Economics*, *17*, 357-390.

Leung, M. and Ong, S. The Predictability of U.S. Market Volatility Index Based on Informational Spillovers from International Markets. Working paper, Indiana University, 1996.

Maberly, E. D., 1986, The Informational Content of the Interday Price Change with Respect to Stock Index Futures, Journal of Futures Markets, 6, 385-395.

McNelis, P., 1996, A Neural Network Analysis of Brazilian Stock Prices: Tequila Effects vs. Pisco Sour Effects, *Journal of Emerging Markets*, 1, 29-44.

Medsker, L., Turban, E., and Trippi, R., Neural Network Fundamentals for Financial Analysts. *Neural Networks in Finance and Investing*, ed. R. Trippi and E. Turban. Chicago: Probes Publishing, 1993.

O'Connor, M., Remus, W., and Griggs, K., 1997, Going Up-Going Down: How Good Are People at Forecasting Trends and Changes in Trends?, Journal of Forecasting, 16, 165-176.

Parzen, E., 1962, On Estimation of a Probability Density Function and Mode. *Annals of Mathematical Statistics*, *33*, 1065-1076.

Refenes, A., Constructive Learning and Its Application to Currency Exchange Rate Forecasting. *Neural Networks in Finance and Investing*, ed. R. Trippi and E. Turban. Chicago: Probes Publishing, 1993.

Rozeff, M., 1984, Dividend Yields are Equity Risk Premiums. *Journal of Portfolio Management*, Fall, 68-75.

Rumelhart, D., and McClelland, J. *Parallel Distributed Processing*. Cambridge, Mass.: MIT Press, 1986.

Schwert, G. and Seguin, P., 1990, Heteroscedasticity in Stock Returns. *Journal of Finance*, *45*, 1129-1155.

Specht, D. Probabilistic Neural Networks for Classification, Mapping, or Associative Memory. *IEEE International Conference on Neural Networks*, 1988.

Specht, D. Probabilistic Neural Networks. *Neural Networks*, 1990, *3*, 109-118.

Wasserman, P. *Advanced Methods in Neural Computing*. New York: Van No strand Reinhold, 1993.

White, J., 1980, Options Values Under Stochastic Volatility: Theory and Empirical Estimates. *Econometrica*, *48*, 817-838.

Wu, Y. And Zhang, H., 1997, Forward Premiums as Unbiased Predictors of Future Currency Depreciation: A Non-Parametric Analysis, Journal of International Money and Finance, 16, 609-623.

APPENDIX

A1. Generalized Method of Moments Estimation

The econometric model used in this study is estimated using the Generalized Method of Moments (GMM). Numerous studies, such as Schwert and Seguin (1989), have presented evidence of heteroscedasticity in stock returns. In addition, non-normality of stock returns has been documented. Blattberg and Gonedes (1974) is one of the many studies which provides evidence concerning the non-normality of stock returns. Hansen (1982) provides a generalized method of moments estimator (GMM) that extends White (1980)'s method to deal with heteroscedasticity and properly deals with serial correlation. This method was first applied by Hansen and Hodrick (1980) to test the predictive power of three-month forward rates measured at monthly intervals. Then, Jorion (1995) applies this method to volatility forecasts. The Hansen-White variance-covariance matrix of estimated coefficients is given by:

$$\Sigma = (X'X)^{-1}\Omega(X'X)^{-1} \qquad \text{(A1)}$$

where $\Omega = E[X'\epsilon\epsilon'X/T]$ is consistently estimated, using the OLS residuals ϵ, by:

$$\hat{\Omega} = \sum_{t} \hat{\epsilon}_t^2 X_t'X_t + \sum_{s}\sum_{t} Q(s,t)\hat{\epsilon}_s\hat{\epsilon}_t(X_t'X_s + X_s'X_t) \qquad \text{(A2)}$$

with *Q(s, t)* defined as an indicator function equal to unity if there is overlap between returns at *s* and t, and zero otherwise. Note that in the case where the residuals are homoscedastic and do not overlap, $E[\epsilon_t^2]=s^2$, and *Q(s, t)* is always zero, so that the covariance matrix collapses to the usual OLS covariance matrix $s^2(X'X)^{-1}$.

A2. Conceptual and Mathematical Foundation of PNN

The construct of the neural network models used in this study is based on the Probabilistic Neural Network (PNN) proposed by Specht (1988, 1990). PNN is fundamentally a multi-layer feedforward classifier. In other words, all network connections are pointing unidirectionally from the current layer toward the subsequent layer. Unlike the backpropagation multi-layer feedforward network (MLFN) widely used in other financial studies, PNN has a much rigid construct. Figure A1 illustrates a typical PNN classifier. The basic network topology consists of four layers -- an input, an output, and two hidden layers. The input layer has one neuron for each variable in the input vector whereas the output layer consists of a set of neurons indicating the class of the input vector. This generalized multivariate configuration can be adapted to the simpler univariate case by introducing only one neuron in the output layer. Each of the network models examined in this study contains a univariate output neuron. The first hidden layer is called the pattern layer and contains one neuron for each training case. The second hidden layer, or the summation layer, is made up of an array of neurons with the number of members equal to the total number of classes. These sizes of the layers can be specified by the user (classification mode) or deduced autonomously during the learning process (autoassociation mode.) The simple PNN construct depicted in Figure A1 represents a model with two input variables (X_1 and X_2), one univariate output (D), two classes, and three training cases for each of the two classes.

Another major distinction of the PNN from the MLFN is how the neurons between the layers are connected. The usual construct of MLFN suggested by many studies contains an "exhaustive" type of connections between any two neurons in adjacent hidden layers, if there is more than one hidden layer. Specifically, each neuron in a particular hidden layer (except the one immediately precedes the output layer) is connected to all neurons in the subsequent layer and there is no grouping of neurons in any hidden layer. On the contrary, PNN groups the neurons in the hidden layers according to the classes they belong. Consequently, a neuron in the pattern layer is only connected to its corresponding class neuron in the summation layer. As shown in Figure A1, each group of three neurons, representing three separate training cases, in the pattern layer is exclusively connected to the summation neuron exemplifying the class of the three cases. There is no connection between the hidden neurons associated with different classes.

PNN is conceptually built on the Bayesian method of classification. Theoretically, unlike the classification scheme which operates the backpropagation MLFN, the Bayesian classification method leads to a convergence. In other words, given enough information, the Bayesian method can classify a new sample with the maximum probability of success (Wasserman (1993)). The principle of a Bayesian classifier rests on the selection of class i with the largest product term in the Bayesian Classification Theorem:

$$\max_i \{h_i \, l_i \, f_i(X)\} \tag{A3}$$

where h_i is the *a priori* probability for class i,

l_i is the loss incurred by misclassifying a sample which truly belongs to class i,

$f_i(\boldsymbol{X})$ is the corresponding probability from the joint density function of class i, and

$\boldsymbol{X}$ is $(x_1, x_2, ..., x_k)$, the input vector to be classified.

When the network is subject to training, a new neuron is added to the existing pattern layer (first hidden layer) for each sample in the training set. Appropriate "class" connections attached to the neuron are then established. The network continues to append pattern neurons to the layer until all training samples have been encountered. Nevertheless, before the actual classification can take place, the Bayesian decision rule requires a prior knowledge of the underlying probability density functions of the observed classes. To avoid possible error caused by misspecification, these density functions are directly estimated from the set of training samples using Parzen's (1962) window approximation method. The PNN used in the current study adopts Cacoullos's (1966) multivariate version of the original Parzen density estimation approach. This extension allows us to generate the joint density functions for a set of k variables. Equation (A4) represents the expression for finding the probability of the joint density function with sample vector $\boldsymbol{X}$. It should be noted that these non-parametric estimations applied in our study is based on Gaussian kernels. Hutchinson et. al. (1994) and Hartman et. al. (1990) suggest that the Gaussian function is a reasonable form for kernel specification and is adaptive to a wide range of situations.

$$f_i(X) = \frac{1}{(2\pi)^{k/2}\sigma^k n_i} \sum_{j=1}^{n_i} e^{\frac{-(X-Y_{ij})'(X-Y_{ij})}{2\sigma^2}} \qquad \text{(A4)}$$

where k is the total number of input variables in the sample vector,

n_i is the number of training samples belongs to class i,

$\boldsymbol{X}$ is the sample vector to be classified,

$\boldsymbol{Y}_{ij}$ is the j^{th} training sample in class i, and

s is a smoothing parameter.

Since the PNN utilizes the Bayesian classification logic in conjunction with the Parzen density estimation technique, the time consuming multiple-epoch training is not necessary. In fact, PNN can be trained in a single cycle if its training time is sufficiently long (Caudill and Butler (1993)). To expedite the training, we allow the network learning process to go through a series of iterative sessions. The learning session is repeated for a sufficient number of times or until convergence is achieved, in that further learning cannot reduce the total network error. Based on our observations in the experiment, the marginal gain in the network performance sharply decreases after a few training cycles have been completed.

After the network has been fully trained, classification begins when an input vector is presented to the pattern layer. Each pattern neuron computes the distance measure between the input and the training case represented by that neuron. The neuron then subjects this measure to its activation function which is the Parzen's density function (Equation A4) estimated in the training process. Subsequently, the summation neuron of each class adds up the output measures from each member pattern neuron. The final output neuron is a threshold discriminator and determines the implied class of the input vector. Specifically, the network will evaluate each of the product terms shown in Equation (A3). If the number of training samples is large and representative, the prior probability h_i for class i can be estimated by the proportion of the number of training samples in class i to the total number of training samples in the training set, that is,

$$h_i = \frac{n_i}{\sum_i n_i} \tag{A5}$$

Substituting Equations (A4) and (A5) into Equation (A6) and assuming that the loss l_i is uniform to all classes, we obtain:

$$\max_i \left\{ \sum_{j=1}^{n_i} e^{\frac{-(X-Y_{ij})'(X-Y_{ij})}{2\sigma^2}} \right\} \tag{A6}$$

Then Equation (A6) indicates the resulting decision criterion for the Bayesian classification performed by PNN. It would be interesting to note that, since the network approximates the activation functions for its neurons, there is no need to declare a specific functional form for activation in the pattern and summation neurons. For this study, we impose an additional criterion to modify the selection of class made by the output neuron. A description of this modification in the output neuron can be found in Section 5.2. The adjustments to the original activation function are outlined in Equations (15) and (16).

Concerning the amount of time devoted for the development of proper network topology, PNN offers a quick and easy solution compared with that offered by the traditional backpropagation MLFN. Although the PNN model involves a more elaborate construct than the backpropagation MLFN, the relatively inflexible hidden layers of PNN requires little effort in the mostly trial-and-error network development process. As mentioned previously, the total number of neurons in the pattern layer is equal to the total number of training cases while the number of neurons in the summation layer is the same as the number of classes for the problem. Since the sizes of the input and output layers are exogenously determined, it is apparent that the iterative procedures for identifying an appropriate number of hidden layers and their respective sizes could be eliminated. However, this special network construct results in a tradeoff between the model simplicity (in development) and the expanding computational effort and memory requirement. Unlike the backpropagation MLFN, the PNN has to perform more computations and consume more memory space due to the large number of neurons in the pattern layer. This deficiency should not constitute a material influence on the network performance given today's computer technology.

The development of ANN model usually encompasses the selection of suitable network topology and the determination of several key parameters associated with training. Among these decision variables are the number of hidden layers, the number of neurons in each of the hidden layers, the number of training cycles in an epoch, the total number of epochs in the complete training session, the learning rate, and the momentum. However, due to the unique characteristics of PNN classifier described earlier, the decision related to the development of network topology and the selection of learning rate and momentum can be disregarded. Instead, the user has to supply a set of smoothing constants, sigmas, to control the shapes of the activation functions in the PNN model. The PNN used in this study contains a self-search routine for finding the best sigma values based on the search range specified by the user.

For each of the PNN models used in our study, there is a total of 68 neurons in the pattern layer and 2 neurons in the summation layer. This configuration represents 68 cases applied to each training session and a total of two classes are allowed for both directions of index returns. Since there are two classes adherent to our designated classification scheme, the neurons in the two hidden layers are divided into two subgroups, one for the negative index returns and another for the positive index returns. Out of the total 68 samples in the training set, 31 samples have negative returns whereas 37 samples indicate positive returns. Hence, there are 31 and 37 pattern neurons associated with Class 1 and Class 2, respectively. These neurons are connected only to their respective summation neurons in the following hidden layer. The weights of the connections pointing from the class neurons to the output neuron signify the respective probabilities of the class affiliation.

Technically, the number of pattern neurons for each class may change to reflect the number of cases associated with each class. For instance, the network construct for the MR3 model used to estimate the second out-of-sample (Period 70) forecast would be different from the construct used to estimate the first out-of-sample (Period 69) forecast. The first construct has 31 and 37 pattern neurons connected to Class 1 and Class 2 summation neurons, respectively, whereas, the second construct contains 30 and 38 pattern neurons for the two classes.

To ensure full training of the network, we designate 40 training sessions in the entire training process. It is interesting to note that convergence (i.e., the total network error cannot be reduced) is mostly attained within 10 sessions in that further training cannot enhance the performance of the network. The reported directions of index returns are the results of the classifications made after 40 sessions have elapsed or convergence has occurred.

Results of the pilot experiment show that the network performance can be enhanced by specifying an independent sigma for each of the elements in the input vector, although the additional computations moderately reduce the speed in training. Since PNN directly estimates the underlying probability density functions from the data set, preprocessing of the raw data, such as rescaling and normalization, is not necessary and will not enhance the performance of the network. Both the exploratory simulation in our pilot study and Leung and Ong (1996) confirm this notion.

Table 1 List of potential economic state input variables and forecasted output variables

Input Variables
TS Term Structure Proxy Three year government bond rate minus the one month risk free rate. The one month risk free rate is the one-month deposit rate at the First Commercial Bank.
TB Short Term Interest Rate One-month deposit rate at the First Commercial Bank.
DS3, DS6, DS12 Lagged Index Returns Continuously compounded lagged three, six, and twelve month annualized excess returns of the Taiwan index respectively.
GC3, GC6, GC12, PC3, PC6, PC12 Consumption Level Continuously compounded lagged three, six, and twelve month annualized growth rates of government consumption and private consumption previous to the period being forecasted.
GNP3, GNP6, GNP12, GDP3, GDP6, GDP12 Gross National and Domestic Products Continuously compounded lagged three, six, and twelve month annualized growth rates of the gross national product and gross domestic product previous to the period being forecasted.
CPI3, CPI6, CPI12, IP3, IP6, IP12 Consumer Price and Production Level Continuously compounded lagged three, six, and twelve month annualized growth rates of the consumer price index and industrial production previous to the period being forecasted.
Output Variables
MR3, MR6, MR12 Returns on Index Continuously compounded three, six, and twelve month annualized excess returns of the Taiwan index. The excess return for a particular time period is defined as the continuously compounded return minus the risk free rate for the corresponding time period.

Table 2. Univariate generalized method of moments regressions

This table presents the results of univariate GMM regressions of various horizon excess returns of the Taiwan index on various constructed macroeconomic state variables. The data for the univariate GMM regressions cover the entire sample period. The t-statistics in parenthesis are heteroscedasticity and autocorrelation consistent. The model estimated is:

$$r_{j,t} = \delta_{j,k,0} + \delta_{j,k,1} instrument_{k,t-1} + \epsilon_{j,t}$$

The dependent variables are: the three, six, and twelve month ahead continuously compounded annualized excess returns of the Taiwan index (MR3, MR6, and MR12). The instruments are the various independent variables described in the study and are observable on or before the last day of the month preceding the month corresponding to the first day of the forecast period.

	TS	TB	DS3	DS6	DS12	GC3	GC6	GC12	PC3	PC6	PC12
MR3	22.3982	-0.2394	0.0760	0.0615	0.2464	-2.6162	-4.3141	-3.7052	0.0835	0.0901	2.9410
	(0.9401)	(-3.4303)*	(0.3908)	(0.2602)	(0.7706)	(-1.3234)	(-1.1525)	(-0.6592)	(0.3270)	(0.2388)	(1.9998)*
MR6	13.5670	-0.2133	0.0206	0.0079	0.0854	-1.3434	-0.3297	-5.9956	0.0168	0.4926	2.5590
	(0.6384)	(-4.1957)*	(0.2268)	(0.0717)	(0.4717)	(-0.8453)	(-0.1232)	(-1.4479)	(0.0783)	(1.2804)	(2.3201)*
MR12	8.6385	-0.1865	0.5540	0.0793	0.1599	-0.1471	-1.8525	-8.3612	0.1383	0.3564	2.0765
	(0.8158)	(-6.3775)*	(1.3343)	(1.2018)	(1.7224)	(-0.1296)	(-0.9772)	(-2.3728)*	(0.9959)	(1.1964)	(2.4491)*

	GNP3	GNP6	GNP12	6DP3	6DP6	6DP12	CPI3	CPI6	CPI12	IP3	IP6	IP12
MR3	-0.0468	0.2000	-7.9804	-0.1323	2.2052	1.3780	-0.0578	2.9569	3.3124	-0.2002	1.4522	5.0814
	(-0.1206)	(0.3113)	(-3.0176)*	(-0.3718)	(1.9751)*	(0.4185)	(-0.0392)	(1.2190)	(1.0406)	(-0.1612)	(0.6719)	(1.4260)
MR6	0.0383	-0.4007	-6.7393	0.1192	1.5118	-1.0307	0.8345	2.6182	2.5011	0.3895	2.1666	4.0535
	(0.1125)	(-0.6447)	(-3.1790)*	(0.5627)	(1.7441)	(-0.3691)	(0.9576)	(1.4463)	(1.0482)	(0.4901)	(1.3130)	(1.5122)
MR12	-0.1472	-0.4357	-5.5598	-0.0118	-0.1948	-3.8667	0.3966	1.1288	3.4233	0.5919	1.4739	4.0326
	(-0.8106)	(-1.1750)	(-4.5487)*	(-0.0798)	(-0.3229)	(-1.6668)	(0.5054)	(0.8750)	(1.6307)	(0.9845)	(1.3565)	(1.9638)*

Table 3 Input Specifications of MR3, MR6, and MR12 Forecasting Models

Excess Return for 3-Month Investment Horizon
TB, GC12, GNP12, GDP6
Excess Return for 6-Month Investment Horizon
TB, GC12, GNP12, GDP6, CPI12
Excess Return for 12-Month Investment Horizon
TB, GC12, GNP12, GDP12, CPI12, IP12

The macroeconomic state variables for each investment horizon represent the input vector (independent variables) to GMM and PNN predictions in both in-sample estimation and out-of-sample forecasting.

Table 4 Comparison of the actual excess return with the directions predicted by each forecasting model over the out-of-sample forecast periods

	Three Month Investment Horizon				Six Month Investment Horizon				Twelve Month Investment Horizon			
Period	**Actual**	**PNN**	**Kalman**	**Rand**	**Actual**	**PNN**	**Kalman**	**Rand**	**Actual**	**PNN**	**Kalman**	**Rand**
69	-2.6193	+	+	+	-0.6027	+	+	+	0.5837	+	+	+
70	0.1347	+	-	+	0.8023	+	-	+	0.7773	+	+	+
71	0.6393	+	+	-	0.7873	+	+	+	0.7863	+	+	+
72	1.4239	+	+	-	1.4111	+	+	+	0.7329	+	+	+
73	1.4799	+	+	+	1.5292	+	+	+	0.7224	+	+	+
74	0.9452	+	+	+	1.5274	+	+	+	0.6441	+	+	+
75	1.4084	+	+	+	1.7800	+	+	-	0.7341	+	+	+
76	1.5885	-	+	+	0.7622	+	+	+	0.5960	+	+	+
77	2.1196	+	+	+	0.7954	+	+	+	0.7369	+	+	+
78	2.1617	+	+	+	0.0646	+	+	+	0.5916	+	+	+
79	-0.0568	+	+	+	-0.0757	+	+	+	0.3681	+	+	+
80	-0.5237	-	+	+	-0.2318	-	+	+	0.2038	+	+	+
81	-2.0274	-	+	+	-0.3044	-	+	+	0.1394	+	+	+
82	-0.0895	-	+	-	0.4373	+	+	+	0.4798	+	+	+
83	0.0652	-	+	-	0.6859	+	+	+	0.2695	+	+	+
84	1.4237	-	+	-	1.1260	+	+	+	0.5788	+	+	+
85	0.9690	-	+	-	0.8207	+	+	-	0.6193	+	+	+
86	1.3115	+	+	+	0.6494	+	+	-	0.4573	+	+	+
87	0.8334	+	+	+	0.5932	+	+	-	0.3228	+	+	+
88	0.6677	-	+	+	0.5226	+	+	+	0.0958	+	+	+
89	-0.0456	-	+	+	-0.1730	+	+	+	-0.3843	+	+	+
90	0.3107	-	-	+	-0.0034	-	-	+	-0.6929	-	-	+
91	0.3473	+	-	+	0.3829	-	-	+	-0.6192	+	-	+
92	-0.3054	+	-	-	0.2308	+	-	+	-1.0993	-	-	+
93	-0.3124	-	-	+	0.0199	+	-	+	-1.4204	-	-	+
94	0.4260	-	-	+	-0.3537	+	-	+	-1.2565	+	-	+
95	0.7745	+	+	-	-0.5984	-	-	-	-0.8596	-	-	+
96	0.3622	+	-	-	-1.3774	-	-	-	-0.8485	-	-	+
97	-1.1234	-	+	+	-1.6143	-	+	+	-1.1922	+	-	+
98	-1.9660	-	-	+	-2.4209	-	-	+	-0.9352	-	-	+
99	-3.1146	-	-	+	-2.8507	-	-	+	-0.8334	-	-	+
100	-2.1027	-	-	-	-2.1493	-	-	-	-0.5456	+	-	+
101	-2.8715	-	-	-	-1.1104	-	-	-	-0.3569	-	-	-
102	-2.5843	-	-	-	-0.3071	-	-	-	0.0380	-	-	-
103	-2.1935	-	-	-	-0.7576	-	-	-	-0.1766	-	-	-
104	0.6555	-	-	-	0.5608	-	-	-	0.1280	-	-	-
105	1.9763	-	-	-	1.1947	-	-	-	0.4924	-	-	-
106	0.6924	+	-	-	1.0706	-	-	-	0.1848	+	-	-
107	0.4812	+	+	+	0.4090	+	+	-	-0.0947	+	+	-
108	0.4276	+	+	+	0.3957	+	+	-	-0.0796	-	+	-
109	1.4679	+	+	+	0.4169	+	+	-	0.1976	-	+	-
110	0.3569	+	+	+	-0.2923	+	+	+	-0.0736	-	+	-
111	0.3837	+	+	+	-0.1966	+	+	+	-0.1632	-	+	-
112	-0.6140	+	+	+	-0.6860	-	+	+	-0.3703	-	+	-
113	-0.9214	-	+	+	-0.5834	-	-	+	-0.3164	-	+	-
114	-0.7568	-	-	+	-0.5398	-	-	+	-0.3380	-	+	+
115	-0.7380	+	-	-	-0.0056	+	-	+	-0.3253	-	+	-
116	-0.2255	-	+	-	0.1626	-	+	-	-0.2334	-	+	+
117	-0.3002	-	-	-	-0.1099	-	-	-	-0.4128	-	+	+
118	0.7507	-	-	-	-0.0311	-	-	-	-0.2761	-	+	+
119	0.5732	+	+	-	-0.0244	-	+	-	-0.2603	-	+	-
120	0.1026	+	+	-	-0.1090	-	+	-	-0.3919	-	+	-
121	-0.7924	+	+	+	-0.6170	-	+	-	-0.5478	-	+	+
122	-0.6018	-	+	+	-0.6020	-	+	+	-0.2369	-	+	-
123	-0.3028	-	+	+	-0.6908	-	+	-	-0.0724	-	+	-
124	-0.4257	-	+	-	-0.4995	-	+	-	-0.0626	-	+	-
125	-0.5871	-	+	-	-0.4760	-	+	-	-0.1297	-	+	-
126	-1.0638	-	+	-	-0.6572	-	+	-	-0.2017	-	+	-
127	-0.5584	-	+	-	-0.4661	-	+	-	-0.1142	-	+	-
128	-0.3499	-	+	-	0.1381	-	+	-	-0.0914	-	+	-

Table 5 Comparison of the predictive strength of each forecasting model

NC denotes the number of times a forecasting model correctly predicts the direction of the index return over the 60 out-of-sample forecast periods. * and ** indicate that the number is statistically different from 30 at a 10% and 5% level of significance, respectively.

	NC		
	3 Month Investment Horizon	**6 Month Investment Horizon**	**12 Month Investment Horizon**
PNN	44**	47**	50**
GMM-Kalman Filter	33	35*	34
Random Walk	32	31	38**

Table 6 Profits realized by various trading strategies over the out-of-sample forecast periods

3 Month Investment Horizon *Initial Investment = $1 per month for three months*															
	0.03 % Commission					**3% Commission**					**6% Commission**				
	BH	RW	KF	ST	MT	BH	RW	KF	ST	MT	BH	RW	KF	ST	MT
Net Gain in Assets ($)	-0.6731	3.9341	4.8558	10.8809	10.9077	-0.7422	2.3061	3.6227	7.5574	7.6426	-0.8121	1.0191	2.5482	4.9438	5.2068
No. of Trades	3	28	17	27	25	3	28	17	27	25	3	28	17	27	25
% Return	-22.44%	131.14%	161.86%	362.70%	363.59%	-24.74%	76.87%	120.67%	251.91%	254.75%	-27.07%	33.97%	84.94%	164.79%	173.56%
6 Month Investment Horizon *Initial Investment = $1 per month for six months*															
	0.03 % Commission					**3% Commission**					**6% Commission**				
	BH	RW	KF	ST	MT	BH	RW	KF	ST	MT	BH	RW	KF	ST	MT
Net Gain in Assets ($)	-0.4276	3.1151	9.4454	15.6901	16.5957	-0.5931	1.9613	7.3122	13.1541	13.5415	-0.7603	0.9233	5.4111	10.5822	10.9314
No. of Trades	6	33	30	25	29	6	33	30	25	29	6	33	30	25	29
% Return	-7.13%	51.92%	157.42%	261.50%	276.60%	-9.89%	32.69%	121.87%	219.24%	225.69%	-12.67%	15.39%	90.18%	180.87%	182.19%
12 Month Investment Horizon *Initial Investment = $1 per month for twelve months*															
	0.03 % Commission					**3% Commission**					**6% Commission**				
	BH	RW	KF	ST	MT	BH	RW	KF	ST	MT	BH	RW	KF	ST	MT
Net Gain in Assets ($)	1.5988	5.2364	21.7005	24.2200	33.8743	1.1948	4.0114	18.7853	21.7753	30.6043	0.7867	2.8348	16.0163	19.4199	27.4719
No. of Trades	12	31	36	26	28	12	31	36	26	28	12	31	36	26	28

% Return	13.32%	43.64%	180.84 %	201.83 %	282.29 %	9.96%	33.43%	156.54 %	181.46 %	255.04 %	6.56%	23.62%	133.47 %	161.83 %	228.93 %

Legend for the investment strategies:

BH: Buy and hold
RW: Random walk
KF: GMM with Kalman filter
ST: PNN with single triggering threshold at 0.5 level
MT: PNN with multiple triggering thresholds at 0.5 and 0.7 levels