Expert Systems with Applications 37 (2010) 7056–7064

Contents lists available at ScienceDirect

Expert Systems with Applications

journal homepage: www.elsevier.com/locate/eswa

Integrating independent component analysis-based denoising scheme with neural network for stock price prediction

Chi-Jie Lu *

Department of Industrial Engineering and Management, Ching Yun University, Jhong-Li, Taiwan, ROC

ARTICLE INFO

Keywords:
Independent component analysis
Neural network
Stock price prediction
Financial forecasting

ABSTRACT

The forecasting of stock price is one of the most challenging tasks in investment/financial decision-making since stock prices/indices are inherently noisy and non-stationary. In this paper, an integrated independent component analysis (ICA)-based denoising scheme with neural network is proposed for stock price prediction. The proposed approach first uses ICA on the forecasting variables to generate the independent components (ICs). After identifying and removing the ICs containing the noise, the rest of the ICs are then used to reconstruct the forecasting variables. The reconstructed forecasting variables will contain less noise information and are served as the input variables of the neural network model to build the forecasting model. The TAIEX closing index and Nikkei 225 opening index are used as illustrative examples to evaluate the performance of the proposed model. Experimental results show that the proposed model outperforms the integrated wavelet denoising technique with BPN model, the BPN model with non-filtered forecasting variables, and a random walk model.

1. Introduction

Stock price prediction is one of the major activities in all private and institution investors. It is an important issue in investment/financial decision-making and is currently receiving considerable attention from the research community. However, it is regarded as one of the most challenging problems due to the fact that stock prices/indices are inherently noisy and non-stationary (Hall, 1994; Li, Li, Zhu, & Ogihara, 2003; Yaser & Atiya, 1996).

There have been many studies using artificial neural networks (ANNs) for stock price prediction (Atsalakis & Valavanis, 2009; Cao & Parry, 2009; Chang, Liu, Lin, Fan, & Ng, 2009; Chavarnakul & Enke, 2008; Enke & Thawornwong, 2005; Hassan, Nath, & Kirley, 2007; Kim, 2006; Tsang et al., 2007; Vellido, Lisboa, & Vaughan, 1999; Yudong & Lenan, 2009; Zhang, Patuwo, & Hu, 1998; Zhu, Wang, Xu, & Li, 2008). A large number of successful applications have shown that ANNs can be useful techniques for stock price forecasting due to their ability to capture subtle functional relationships among the empirical data even though the underlying relationships are unknown or hard to describe (Atsalakis & Valavanis, 2009; Vellido et al., 1999; Zhang et al., 1998). Unlike traditional statistical models, known as Box–Jenkins ARIMA (Box & Jenkins, 1970), ANNs are data-driven and non-parametric models. They do not require strong model assumptions and can map any nonlinear function without a priori assumption about the properties of the data (Chauvin & Rumelhart, 1995; Haykin, 1999; McNelis, 2004). The most popular neural network training algorithm for financial forecasting is the backpropagation neural network (BPN) (Atsalakis & Valavanis, 2009; Cao & Parry, 2009; Chang et al., 2009; Lee & Chen, 2002; Lee & Chiu, 2002; McNelis, 2004; Vellido et al., 1999; Yudong & Lenan, 2009; Zhang et al., 1998), which has a simple architecture but a powerful problem-solving ability. It motivates this study of using BPN for stock price prediction.

When using BPN for stock price forecasting, one of the key problems is the inherent noise of the stock prices/indices. Learning observations with noise without paying attention may lead to fitting those unwanted data and may torture the approximation function. This will result in the loss of generalization capability in the testing phase. Moreover, the noise in the data could lead to over-fitting or under-fitting problems (Li et al., 2003; McNelis, 2004; Yu, Wang, & Jiang, 1996; Zhao, Chen, & Hu, 2004). Therefore, detecting and removing the noise are important but difficult tasks when building a BPN forecasting model. Few studies have been proposed to deflate the influence of noisy data and enhance the robust capability of BPN (Yu et al., 1996; Zhao et al., 2004). However, the existing methods would either involve extensive computation or use additional parameters for BPN algorithm. When the parameters are not properly chosen, the final results may be affected by its parameters. Moreover, the selection of parameters is not straightforward. To alleviate the influence of noise, in this research, an independent component analysis (ICA)-based denoising scheme

* Tel.: +886 3 4581196x6712; fax: +886 3 4683298.
E-mail addresses: jerrylu@cyu.edu.tw, Chijie.lu@gmail.com.

0957-4174/$ - see front matter
doi:10.1016/j.eswa.2010.03.012

is proposed and integrated with BPN for building a stock price prediction model (called ICA–BPN model).

ICA is a novel statistical signal process technique used to find independent sources given only observed data that are mixtures of the unknown sources, without any prior knowledge of the mixing mechanisms (Cichocki & Amari, 2002; Hyvärinen, Karhunen, & Oja, 2001; Lee, 1998). It is aimed at extracting the hidden information from the observed multivariate statistical data where no relevant signal mixture mechanisms are available. This hidden information is called the independent components (ICs) of the observable data. The noise information of financial time series data usually cannot be directly obtained from the observed data. It might be the latent information of the financial time series. Thus, ICA can be used to detect and remove the noise via the identification of the ICs of the observed financial time series data, and further improve the performance of the BPN forecasting model.

The basic ICA model has been widely applied in signal processing, face recognition and feature extraction (Beckmann & Smith, 2004; Déniz, Castrillón, & Hernández, 2003; James & Gibson, 2003; Jang, Lee, & Oh, 2002; Kim, Kim, Hwang, & Kittler, 2004). However, there are still few applications using ICA in financial time series forecasting (Górriz, Puntonet, Salmerón, & Lang, 2003; Malaroiu, Kiviluoto, & Oja, 2000; Popescu, 2003). Back and Weigend (1997) used ICA to exact the features of the daily returns of the 28 largest Japanese stocks. The results showed that the dominant ICs can reveal more underlying structure and information of the stock prices than principal component analysis. Kiviluoto and Oja (1998) employed ICA to find the fundamental factors affecting the cash flow of the 40 stores belong to the same retail chain. They found that the cash flow of the retail stores was mainly affected by holidays, seasons and competitors' strategies. Oja, Kiviluoto, and Malaroiu (2000) applied ICA in foreign exchange rate time series prediction. They firstly used ICA to estimate independent components and mixing matrix from the observed time series data and filtered the independent components to reduce the effects of noise through the linear and nonlinear smoothing techniques. Then, the autoregression (AR) modeling was employed to predict the smoothed independent components. Finally, they combined the predictions of each smoothed IC by using mixing matrix and thus obtained the predications for the original observed time series.

The proposed ICA–BPN approach first uses ICA on the forecasting variables to estimate the independent components. Since the stock prices/indices are inherently noisy, at least one IC can be used to represent noise information of the data. After identifying and removing the ICs containing the noise, the rest of the ICs are then used to reconstruct the forecasting variables. The reconstructed forecasting variables will contain less noise information and serve as the input variables of the BPN for building the forecasting model. In order to evaluate the performance of the proposed approach, the Taiwan Stock Exchange Capitalization Weighted Stock Index (TAIEX) closing cash index and Nikkei 225 opening cash index are used as the illustrative examples.

The rest of this paper is organized as follows. Section 2 gives brief overviews of ICA and neural networks. The proposed forecasting model is thoroughly described in Section 3. Section 4 presents the experimental results from the datasets including the TAIEX closing cash index and Nikkei 225 opening cash index. The paper is concluded in Section 5.

2. Research methodology

2.1. Independent component analysis

Let $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_m]^T$ be a multivariate data matrix of size $m \times n, m \leqslant n$, consisting of observed mixture signals $\mathbf{x}_i$ of size $1 \times n, i = 1, 2, \ldots, m$. In the basic ICA model, the matrix $\mathbf{X}$ can be modeled as (Hyvärinen & Oja, 2000)

$$\mathbf{X} = \mathbf{AS} = \sum_{i=1}^{m} \mathbf{a}_i \mathbf{s}_i \tag{1}$$

where $\mathbf{a}_i$ is the ith column of the $m \times m$ unknown mixing matrix $\mathbf{A}$; $\mathbf{s}_i$ is the ith row of the $m \times n$ source matrix $\mathbf{S}$. The vectors $\mathbf{s}_i$ are latent source signals that cannot be directly observed from the mixture signals $\mathbf{x}_i$. The ICA model aims at finding an $m \times m$ de-mixing matrix $\mathbf{W}$ such that

$$\mathbf{Y} = [\mathbf{y}_i] = \mathbf{WX}, \tag{2}$$

where $\mathbf{y}_i$ is the ith row of the matrix $\mathbf{Y}$, $i = 1, 2, \ldots, m$. The vectors $\mathbf{y}_i$ must be as statistically independent as possible, and are called independent components (ICs). When de-mixing matrix $\mathbf{W}$ is the inverse of mixing matrix $\mathbf{A}$, i.e. $\mathbf{W} = \mathbf{A}^{-1}$, ICs ($\mathbf{y}_i$) can be used to estimate the latent source signals $\mathbf{s}_i$.

The ICA modeling is formulated as an optimization problem by setting up the measure of the independence of ICs as an objective function and using some optimization techniques for solving the de-mixing matrix $\mathbf{W}$. Several existing algorithms can be used to perform ICA modeling (Bell & Sejnowski, 1995; David & Sanchez, 2002; Hyvärinen et al., 2001). In general, the ICs are obtained by using the de-mixing matrix $\mathbf{W}$ to multiply the matrix $\mathbf{X}$, i.e. $\mathbf{Y} = \mathbf{WX}$. The de-mixing matrix $\mathbf{W}$ can be determined using an unsupervised learning algorithm with the objective of maximizing the statistical independence of ICs. The ICs with non-Gaussian distributions imply the statistical independence (Hyvärinen et al., 2001), and the non-Gaussianity of the ICs can be measured by the negentropy:

$$J(\mathbf{y}) = H(\mathbf{y}_{gauss}) - H(\mathbf{y}) \tag{3}$$

where $\mathbf{y}_{gauss}$ is a Gaussian random vector having the same covariance matrix as $\mathbf{y}$. H is the entropy of a random vector $\mathbf{y}$ with density $p(\mathbf{y})$ defined as $H(\mathbf{y}) = -\int p(\mathbf{y}) \log p(\mathbf{y}) d\mathbf{y}$.

The negentropy is always non-negative and is zero if and only if $\mathbf{y}$ has a Gaussian distribution. Since the problem in using negentropy is computationally very difficult, an approximation of negentropy is proposed as follows (Hyvärinen et al., 2001):

$$J(y) \approx [E\{G(y)\} - E\{G(v)\}]^2 \tag{4}$$

where v is a Gaussian variable of zero mean and unit variance, and y is a random variable with zero mean and unit variance. G is a non-quadratic function, and is given by $G(y) = \exp(-y^2/2)$ in this study. The FastICA algorithm proposed by Hyvärinen et al. (2001) is adopted in this paper to solve for the de-mixing matrix $\mathbf{W}$.

2.2. Neural networks

A neural network is a massively parallel system comprising highly interconnected, interacting processing elements (often termed units, nodes or neurons) that are based on neurobiological models. Neural networks process information through the interactions of a large number of simple processing elements. Knowledge is not stored within individual processing units but is represented by the weight between units (Cheng & Titterington, 1994). Each piece of knowledge is a pattern of activity spread among many processing elements, and each processing element can be involved in the partial representation of many pieces of information. Owing to its associated memory characteristic and its generalization capability, neural networks are found to be useful in modeling non-stationary processes (Cheng & Titterington, 1994; Haykin, 1999; Stern, 1996).

Networks architecture is organization of nodes and the types of connections permitted. The network consists of a number of nodes connected by links. The nodes in the neural network can be divided into three-layers: input, output and one or more hidden layers. The nodes in the input layer receive input signals from an external source and the nodes in the output layer provide the target output signals. Any layers between input and output layers are called hidden layers. Various network architectures and learning algorithms have been developed. The backpropagation neural network (BPN) is feedforward network and is probably the most commonly used class of neural network in financial time series forecasting and business (Atsalakis & Valavanis, 2009; Vellido et al., 1999; Zhang et al., 1998). A three-layer BPN is used in this study (as seen in Fig. 1).

BPN is essentially a gradient steepest descent training algorithm. For the gradient descent algorithm, the step size, called the learning rate, must be specified first. The learning rate is crucial for BPN since smaller learning rates tend to slow down the learning process before convergence while larger ones may cause network oscillation and unable to converge. As to the issue of determining the appropriate network topology (the number of layers, the number of nodes in each layer, and the appropriate learning rates), please refer to Chauvin and Rumelhart (1995) and Haykin (1999) for more details. Detailed descriptions of using neural networks in the applications of forecasting and business can be found in McNelis (2004), Vellido et al. (1999), and Zhang et al. (1998).

3. Proposed ICA–BPN forecasting model

In the proposed ICA–BPN prediction model, ICA is first applied to filter out the noise contained in forecasting variables. The filtered forecasting variables are then used in BPN for constructing a forecasting model. When using ICA for denoising, the basic ICA model is first utilized on the mixture matrix $\mathbf{X}$ of size $m \times n$ combined with m forecasting variables ($\mathbf{x}_i$) of size $1 \times n$ for estimating a de-mixing matrix ($\mathbf{W}$) of size $m \times m$ and independent components ($\mathbf{y}_i$) of size $1 \times n$. Since the noise of stock prices may contain the least information about the trend or main feature of the data, the ICs which cannot capture the main feature of stock indices are used to represent the noise of the data. To find the IC representing the noise, the testing-and-acceptance (*TnA*) approach, proposed by Cheung and Xu (2001) using relative hamming distance (RHD) reconstruction error as its index is adapted to order the ICs.

An example is used for illustrating the concept of the *TnA* method. Fig. 2 shows six financial time series data, each of size 1×781. They can be combined as a mixture matrix $\mathbf{X}$ of size 6×781. After using ICA to the matrix $\mathbf{X}$, a de-mixing matrix $\mathbf{W}$ of size 6×6 and four ICs, each of size 1×781, can be estimated. The profiles of

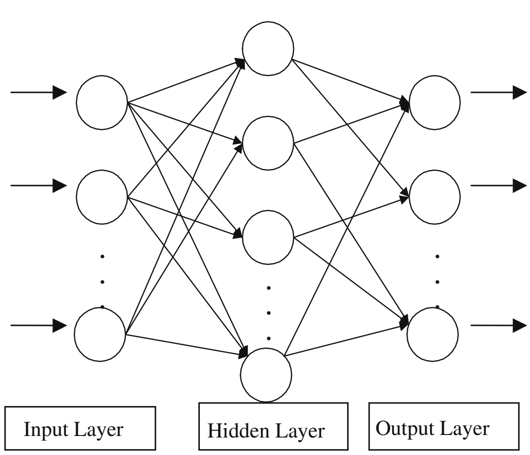


Fig. 1. A three-layer backpropagation neural network.

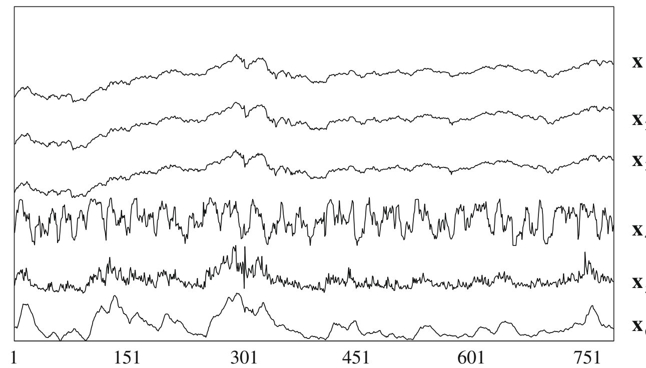


Fig. 2. Six financial time series data of size 1×781, respectively.

those six ICs are shown in Fig. 3. It can be seen from Fig. 3 that each IC can represent different features of the original time series data in Fig. 2. For evaluating the performance of the six ICs on capturing the main feature of the data in Fig. 2, the *TnA* approach is then used.

Consider a set of m ICs, i.e. $m = 6$ in this example. In the first iteration of the *TnA* algorithm, each IC is assumed as the last one in the ordering and is excluded in reconstructing the mixture matrix. Let $\mathbf{y}_k$ be the last IC in the ordering, the reconstructed mixture matrix $\mathbf{X}^R$ can be obtained by using the following equation,

$$\mathbf{X}^R = \sum_{i=1, i \neq k}^{m} \mathbf{a}_i \mathbf{y}_i, \quad 1 \leqslant k \leqslant m \tag{5}$$

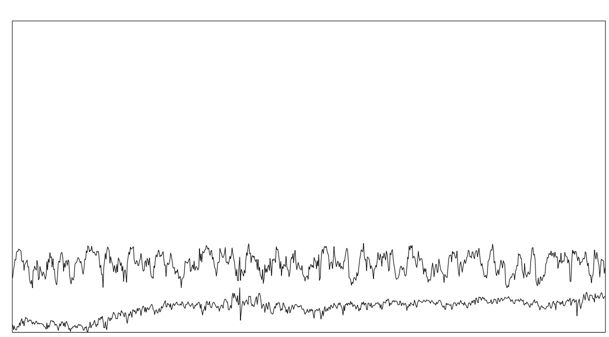
where $\mathbf{X}^R = \left[\mathbf{x}_1^R, \mathbf{x}_2^R, \ldots, \mathbf{x}_m^R\right]^T$ is reconstructed mixture matrix of size $m \times n$, $\mathbf{x}_i^R$ is the ith reconstructed forecasting variable, $\mathbf{a}_i$ is the ith column vector of mixing matrix $\mathbf{A}$, $\mathbf{A} = \mathbf{W}^{-1}$, and $\mathbf{y}_i$ is the ith IC. After obtaining the reconstructed matrices considering the different IC as the last IC in the ordering, the RHD reconstruction error between each reconstructed matrix $\mathbf{X}^R$ and original mixture matrix $\mathbf{X}$ can be computed using the following equation (Cheung & Xu, 2001),

$$RHD = \sum_{i=1}^{m} \left(\frac{1}{n-1} \sum_{t=1}^{n-1} [R_i(t) - \widehat{R}_i(t)]^2 \right) \tag{6}$$

where $R_i = sign[\mathbf{x}_i(t+1) - \mathbf{x}_i(t)]$; $\widehat{R}_i = sign[\mathbf{x}_i^R(t+1) - \mathbf{x}_i^R(t)]$;

$$sign(r) = \begin{cases} 1 & \text{if } r > 0 \\ 0 & \text{if } r = 0 \\ -1 & \text{otherwise} \end{cases}.$$

The RHD reconstruction error can evaluate the similarity between time series. The smaller RHD value represents the higher similarity between time series. The RHD value of identical time series is zero, whereas the RHD value of totally different time series is four. More precisely, the RHD value can be used to evaluate the



similarity between the original forecasting variables and their corresponding reconstructed forecasting variables. That is, if the RHD value of the original forecasting variables and their reconstructed forecasting variables is close to zero, the main feature of the original variables and the reconstructed variables are similar. The IC used to generate the reconstructed forecasting variables will contain the main feature information of the original forecasting variables. Conversely, the RHD value between the original forecasting variables and their reconstructed variables is far from zero. The corresponding IC used for reconstruction will contain less main information of the forecasting variables. The TnA approach is an approximate search algorithm. For more detail information about TnA algorithm, please refer to Cheung and Xu (2001).

Table 1 illustrates the RHD reconstruction error of every iteration of the *TnA* approach used for ordering the ICs in Fig. 3. It can be seen from Table 1 that the reconstructed matrix $\mathbf{X}^R$ excluded IC4 has the smallest RHD value in the first iteration. This indicates that IC4 contributes the least information in the reconstruction of original data. It contains less information about the main feature of the original forecasting variables than that of the rest ICs. Thus, IC4 is selected as the last one IC in the ordering and removed from this sorting process before next iterations. In the next iterations, the *TnA* algorithm repeats the same operations using Eqs. (5) and (6) on the remaining $m - 1$ independent components, i.e. five ICs in this example, and select the second-last component, ..., and so forth. From Table 1, it can be observed that the second-last IC in the ordering is IC2 since the reconstruction set excluding IC2 has the smallest RHD value in the second iteration. According to Table 1, the six ICs can be ordered as follows: IC5, IC6, IC3, IC1, IC2 and IC4. The first one IC, i.e. IC5, contains the most information about the main feature of the time series data (i.e. forecasting variables) in the Fig. 2. Conversely, the least one IC, i.e. IC4, can be used to represent the noise of the data in Fig. 2 since it includes the least main information of the data.

Although the first one of the sorted IC contains most information of the data, the remaining ICs still involve different levels of the main information according to their sorted order. Thus, for fully capturing the main features, the de-noised forecasting variables can be obtained by using all ICs excluding the IC representing the noise. That is, in this example, the filtered forecasting variables are gained by using IC1, IC2, IC3, IC5 and IC6 for reconstruction. Fig. 4 shows the series data of the $\mathbf{x}_1$ in Fig. 2 and its two reconstructed series data, respectively using IC1, IC2, IC3, IC5 and IC6, and only IC4. It can be seen from Fig. 4 that the trend and shape of the reconstructed series data using IC1, IC2, IC3, IC5 and IC6 are very similar to its original series data, i.e. $\mathbf{x}_1$. Conversely, the reconstructed series data using only IC4 is very different from the original series data.

After obtaining the de-noised forecasting variables, they are then used in building the BPN forecasting model. Since one hidden layer network is sufficient to model any complex system with desired accuracy (Chauvin & Rumelhart, 1995) the designed BPN model in this study will have only one hidden layer. The performance of BPN is mainly affected by the setting of network topology, i.e. the number of nodes in each layer and learning rates. There are no general rules for the choice of network topology. The selection is usually based on the trial-and-error (or called cross-validation) method. In this study, the optimal network topology of the BPN model is determined by the trial-and-error method.

4. Empirical study

4.1. Datasets and performance criteria

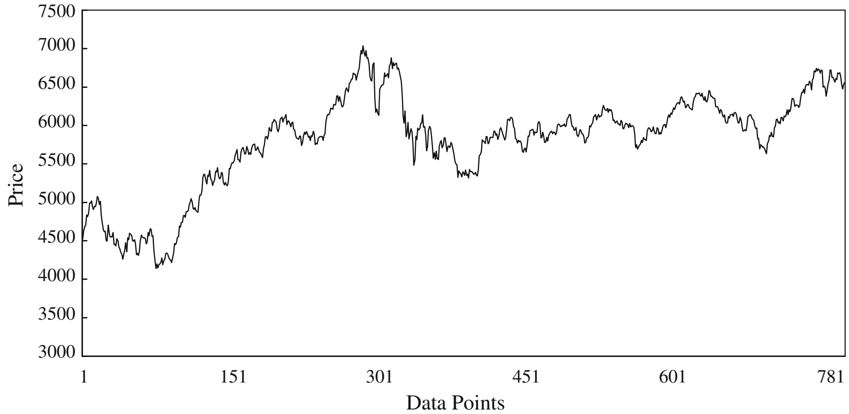


Fig. 5. The daily TAIEX closing cash prices from January 2, 2003 to February 27, 2006.

sample and the remaining 235 data points (30.10% of the total sample points) are used as testing sample.

In forecasting Nikkei 225 opening cash index, the Nikkei 225 index futures prices are used as forecasting variables since the futures price changes lead price changes of the cash market (Lee & Chen, 2002; Lee & Chiu, 2002). Using the leading futures as forecasting variables should contribute to the success in increasing the forecasting accuracy. There are three Nikkei 255 index futures contracts traded on Singapore Exchange-Derivative Trading Limited (SGX-DT), Osaka Securities Exchange (OSE) and Chicago Mercantile Exchange (CME) markets. The previous day's cash market closing index is also an important variable for predicting the cash market opening price (Lee & Chen, 2002; Lee & Chiu, 2002). Therefore, four forecasting variables are used for predicting the Nikkei 225 opening cash index. The daily data of futures and cash prices from February 2, 2004 to February 29, 2008 of the Nikkei 225 cash index provided by Bloomberg are collected in this study. There are totally 1000 data points in the dataset and the daily Nikkei 225 opening cash prices are shown in Fig. 6. The first 700 data points (70.00% of the total sample points) are used as the training sample while the remaining 300 data points (30.00% of the total sample points) are used as the testing sample.

The prediction performance is evaluated using the following performance measures, namely, the root mean square error (RMSE), mean absolute percentage error (MAPE) and directional accuracy (DA). The definitions of these criteria can be found in Table 2. RMSE and MAPE are measures of the deviation between actual and predicted values. The smaller values of RMSE and MAPE, the closer are the predicted time series values to that of the actual value. They can be used to evaluate the prediction error. DA provides the correctness of the predicted direction of the cash index in terms of percentage. DA can be utilized to evaluate the prediction accuracy.

4.2. Forecasting results

The forecasting results of the proposed ICA–BPN model are compared to the BPN model using non-filtered forecasting variables (called single BPN model), the random work model which simply uses the previous day's price to predict today's price, and the integrated wavelet denoising technique with BPN model (called Wavelet-BPN model).

Wavelet denoising techniques based on discrete wavelet transform is a well-established noise removal technique and has been widely applied in the fields of image processing, signal processing and time series analysis (Chang, Yu, & Vetterli, 2000; Donoho & Johnstone, 1998; Hussain, Reaz, Mohd-Yasin, & Ibrahimy, 2009; Li et al., 2003; Ramsey, 2002). In the Wavelet-BPN model, the wavelet denoising techniques based on hard thresholding method is first used to filter out the noise from the forecasting variables. The BPN then utilizes the filtered forecasting variables to build the forecasting model. When using wavelet denoising techniques, two parameters must be settled. They are wavelet basis function and levels of decompose (scales). Most researches of using wavelet transform in financial time series analysis are modeling with Haar (Db1) basis function (Assume, Campbell, & Murtage, 1998; Li et al., 2003; Zheng, Starck, Campbell, & Murtagh, 1999). And, after pretesting different levels with Haar basis function, level 3 seems to

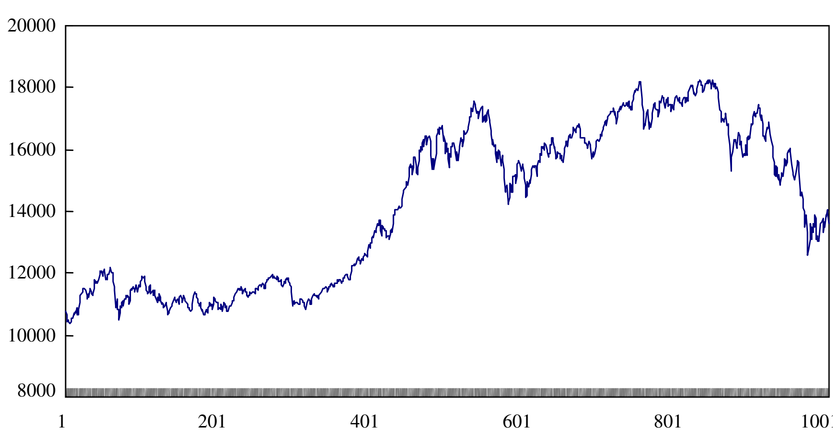


Fig. 6. The daily Nikkei 225 closing cash prices from February 2, 2004 to February 29, 2008.

Table 2
Performance measures and their definitions.

Metrics	Calculation
RMSE	$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(T_i - P_i)^2}{n}}$
MAPE	$MAPE = \frac{\sum_{i=1}^{N}\left\lvert\frac{T_i - P_i}{T_i}\right\rvert}{N}$
DA	$DA = \frac{100}{n}\sum_{i=1}^{n} d_i$, where $d_i = \begin{cases} 1 & (P_i - P_{i-1})(T_i - T_{i-1}) \geqslant 0 \\ 0 & \text{otherwise} \end{cases}$

*Note that T and P represent the actual and predicted value, respectively, n is total number of data points.

work best with most sets of the data of this study. For more detail information about wavelet transform and wavelet denoising technique, please refer to Rao and Bopardikar (1998), Assume et al. (1998), Zheng et al. (1999), Gonzalez and Woods (2002) and Li et al. (2003).

For building BPN forecasting model, the neural network toolbox of MATLAB software is adapted in this study. The wavelet toolbox of MATLAB software is used for performing wavelet denoising. The original data are scaled into the range of [−1.0, 1.0] when building BPN forecasting model. The linear scaling ensures the large value input variables do not overwhelm smaller value inputs, then helps to reduce prediction errors.

In the modeling of the proposed ICA–BPN model, the noise of six forecasting variables should be removed first using the ICA approach. As the six forecasting variables are the time series discussed and expressed in Fig. 2, the noise removing process and results have been discussed in Section 3. After using ICA to filter out the noise of the six forecasting variables, the de-noised forecasting variables are then used for building the BPN forecasting model. The input layer has six nodes as six forecasting variables are used. Since there are no general rules for the choice of the number of the hidden layer, the number of hidden nodes to be tested was set to 11, 12, 13 and 14. And the network has only one output node, the forecasted closing cash price index. As lower learning rates tended to give the best network results (Chauvin & Rumelhart, 1995), learning rates 0.01, 0.02, 0.03, 0.04 and 0.05 are tested during the training process.

In the modeling of the single BPN model and Wavelet-BPN model, like the ICA–BPN model, the input layer has six nodes. In single BPN model, the number of hidden nodes to be tested was 9, 10, 11 and 12. The tested number of hidden nodes for Wavelet-BPN model was 11, 12, 13 and 14. Again the network has only one output node, i.e. the forecasting closing cash price index, and learning rates 0.01, 0.02, 0.03, 0.04 and 0.05 are tested during the training process. The convergence criteria used for training the ICA–BPN, Wavelet-BPN and single BPN models are a root mean squared error (RMSE) less than or equal to 0.0001 or maximum of 1000 iterations. The network topology with the minimum testing RMSE is considered as the optimal network.

The testing results of the ICA–BPN, single BPN and Wavelet-BPN models with combinations of different hidden nodes and learning rates are summarized in Tables 3–5. From Table 2, it can be observed that the {6-13-1} topology with a learning rate of 0.04 gives the best forecasting result (minimum testing RMSE) and hence is the best topology setup for the proposed ICA–BPN model in forecasting TAIEX closing cash index. Here, {6-13-1} represents the six nodes in the input layer, 13 nodes in the hidden layer and one node in the output layer. Table 4 shows that the {6-10-1} topology with a learning rate of 0.03 gives the optimal topology setting for the single BPN forecasting model. As depicted in Table 5, the {6-13-1} topology with a learning rate of 0.02 is the best topology setup for the Wavelet-BPN model.

Table 3
Model selection results of the proposed ICA–BPN forecasting model.

Number of nodes in the hidden layer	Learning rate	Training RMSE	Testing RMSE
11	0.01	0.015570	0.011316
	0.02	0.015500	0.011198
	0.03	0.015044	0.011179
	0.04	0.014062	0.010863
	0.05	0.014044	0.010847
12	0.01	0.015035	0.011002
	0.02	0.014592	0.010875
	0.03	0.014200	0.011033
	0.04	0.014592	0.010875
	0.05	0.014090	0.010817
13	0.01	0.015569	0.011302
	0.02	0.015036	0.010976
	0.03	0.014053	0.010853
	0.04	**0.014227**	**0.010530**
	0.05	0.014053	0.010853
14	0.01	0.015547	0.011318
	0.02	0.015479	0.011246
	0.03	0.015352	0.011169
	0.04	0.014098	0.010740
	0.05	0.014087	0.010720

Table 4
Model selection results of the single BPN forecasting model.

Number of nodes in the hidden layer	Learning rate	Training RMSE	Testing RMSE
9	0.01	0.015939	0.011772
	0.02	0.015527	0.011281
	0.03	0.015403	0.011246
	0.04	0.015392	0.011260
	0.05	0.015939	0.011772
10	0.01	0.015838	0.011598
	0.02	0.015537	0.011295
	0.03	**0.015439**	**0.011213**
	0.04	0.015403	0.011395
	0.05	0.015838	0.011598
11	0.01	0.015885	0.011669
	0.02	0.015528	0.011251
	0.03	0.015409	0.011309
	0.04	0.015388	0.011372
	0.05	0.015881	0.011663
12	0.01	0.015516	0.011325
	0.02	0.015384	0.011370
	0.03	0.015434	0.011336
	0.04	0.015885	0.011669
	0.05	0.015528	0.011251

To examine the convergence characteristics of the proposed ICA–BPN, single BPN and Wavelet-BPN models, the RMSE during the testing process for the {6-13-1}, {6-10-1} and {6-13-1} networks are displayed in Figs. 7–9, respectively. The excellent convergence characteristics of the three models can easily be observed.

The TAIEX closing cash price index forecast results using the proposed ICA–BPN, Wavelet-BPN, single BPN and random walk models are computed and listed in Table 6. Table 6 depicts that the RMSE and MAPE of the proposed ICA–BPN model are, respectively, 40.54 and 0.49%. It can be observed that these values are smaller than those of the Wavelet-BPN, single BPN and random walk models. It indicates that there is a smaller deviation between the actual and predicted values using the proposed ICA–BPN model. Moreover, compared to the Wavelet-BPN, single BPN and random walk models, the ICA–BPN model has the highest DA ratio which is 61.56 DA provides a good measure of the consistency in

Table 5
Model selection results of the Wavelet-BPN forecasting model.

Number of nodes in the hidden layer	Learning rate	Training RMSE	Testing RMSE
11	0.01	0.017100	0.012707
	0.02	0.016716	0.012380
	0.03	0.016671	0.012510
	0.04	0.016665	0.012681
	0.05	0.016626	0.012680
12	0.01	0.017132	0.012531
	0.02	0.016768	0.012272
	0.03	0.016586	0.012458
	0.04	0.016933	0.012582
	0.05	0.016723	0.012566
13	0.01	0.017150	0.012508
	0.02	**0.016718**	**0.012262**
	0.03	0.016729	0.012297
	0.04	0.017088	0.012706
	0.05	0.016741	0.012487
14	0.01	0.017027	0.012647
	0.02	0.016729	0.012288
	0.03	0.019040	0.014788
	0.04	0.019442	0.014478
	0.05	0.019636	0.014834

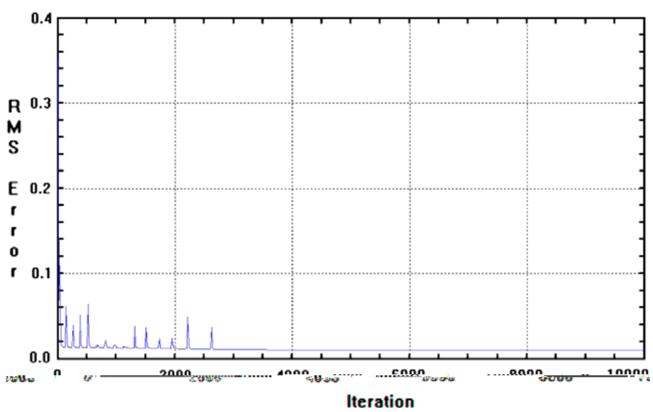


Fig. 7. RMSE history in the testing process for the {6-13-1} network of the proposed ICA–BPN model.

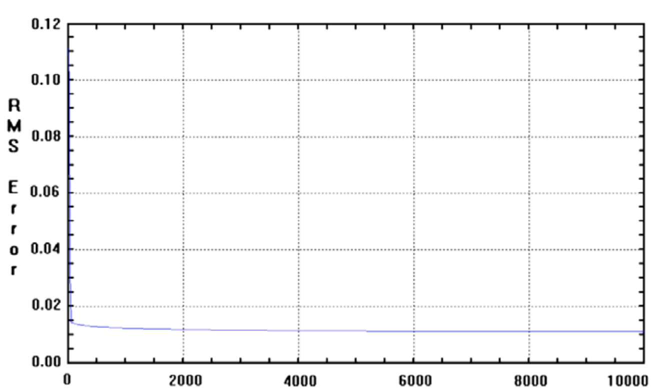


Fig. 8. RMSE history in the testing process for the {6-10-1} network of the single BPN model.

prediction of the price direction. Thus, the proposed ICA–BPN model provides a better forecasting result than the Wavelet-BPN, single BPN and random walk models in terms of prediction error and prediction accuracy.

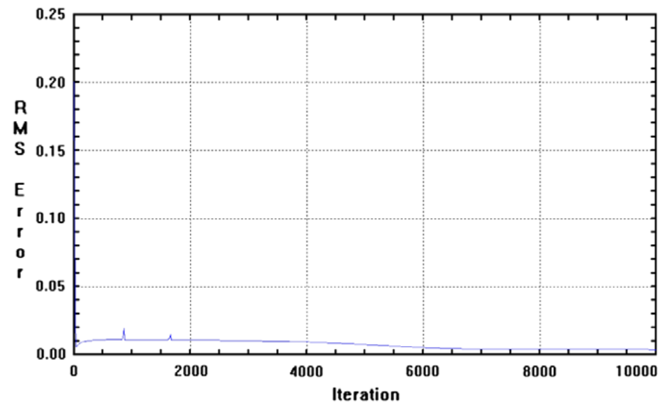


Fig. 9. RMSE history in the testing process for the {6-13-1} network of the Wavelet-BPN model.

Table 6
The TAIEX closing cash prices forecasting results using the proposed ICA–BPN, Wavelet-BPN, single BPN and random walk models.

Metrics	RMSE	MAPE (%)	DA (%)
Models			
ICA–BPN model	40.54	0.49	61.56
Wavelet-BPN model	47.21	0.56	56.98
Single BPN model	46.37	0.55	56.84
Random walk	53.21	0.65	46.15

Table 7
The Nikkei 225 opening cash prices forecasting results using the proposed ICA–BPN, Wavelet-BPN, single BPN and random walk models.

Metrics	RMSE	MAPE (%)	DS (%)
Models			
ICA–BPN model	43.54	0.52	86.25
Wavelet-BPN model	50.72	0.56	80.62
Single BPN model	51.32	0.57	78.51
Random walk	215.38	0.99	46.67

The proposed ICA–BPN method also performs well in forecasting the Nikkei 225 opening cash prices. Table 7 summarizes the Nikkei 225 opening cash prices forecasting results using the ICA–BPN, Wavelet-BPN, single BPN and random walk models. It can also be observed from Table 7 that the proposed ICA–BPN model has the smallest RMSE and MAPE values and the highest DA value in comparison with the Wavelet-BPN, single BPN and random walk models. Thus, the proposed method can produce lower prediction errors and higher prediction accuracy on the direction of change in price and outperforms the Wavelet-BPN, single BPN and random walk models in forecasting of the Nikkei 225 opening cash prices.

4.3. Robustness evaluation

To evaluate the robustness of the proposed ICA–BPN method, the performance of the Wavelet-BPN, single BPN, random walk and proposed models was tested using different ratios of training and testing sample sizes. The testing plan is based on the relative ratio of the size of the training dataset size to complete dataset size. In this section, four relative ratios, 60, 70, 80, and 90% are considered. The prediction results for the TAIEX closing cash index and Nikkei 225 opening cash index by the four methods

Table 8
Robustness evaluation of ICA–BPN, Wavelet-BPN, single BPN and random walk models by different training and testing sample sizes.

Relative ratio (%)	Models	TAIEX		Nikkei 225	
		Testing data MAPE (%)	Testing data DA (%)	Testing data MAPE (%)	Testing data DA (%)
60	ICA–BPN	**0.50**	**62.86**	**0.59**	**81.81**
	Wavelet-BPN	0.61	57.26	0.70	72.45
	Single BPN	0.55	56.83	0.68	73.42
	Random walk	0.65	47.58	0.94	50.57
70	ICA–BPN	**0.49**	**61.56**	**0.52**	**86.25**
	Wavelet-BPN	0.56	56.98	0.56	80.62
	Single BPN	0.55	56.84	0.57	78.51
	Random walk	0.65	46.15	0.99	46.67
80	ICA–BPN	**0.54**	**64.16**	**0.49**	**87.79**
	Wavelet-BPN	0.64	60.03	0.56	80.51
	Single BPN	0.67	57.66	0.56	79.32
	Random walk	0.68	47.09	1.12	48.00
90	ICA–BPN	**0.53**	**67.98**	**0.41**	**90.17**
	Wavelet-BPN	0.62	61.97	0.48	83.08
	Single BPN	0.63	61.06	0.51	83.52
	Random walk	0.75	37.66	1.42	49.00

are summarized in Table 8 in terms of two criteria, MAPE and directional accuracy (DA).

Based on the findings in Fig. 8, it can be observed that the proposed ICA–BPN method outperforms the other benchmarking tools under all four different ratios in terms of the MAPE and DA criteria. It can produce the lowest prediction error and the highest prediction accuracy under all relative ratios of the two testing datasets. It therefore indicates that the proposed ICA–BPN model indeed provides better forecasting accuracy than the other three approaches. The proposed method can effectively detect and remove the noise from stock price data and improve the forecasting performance of BPN.

5. Conclusions

This paper has presented an integrated ICA-based denoising scheme with neural network model for stock price prediction. The proposed method first uses ICA-based on signal reconstruction criterion to remove the noise from forecasting variables since the stock price data is inherently noise. The noise in the data could lead to the over-fitting or under-fitting problem. The filtered forecasting variables containing less noise information are then used in BPN for building prediction model. The TAIEX closing cash index and Nikkei 225 opening cash index are used in this study for evaluating the performance of the proposed method. This study compared the proposed method with the Wavelet-BPN, single BPN and random walk models using prediction error and prediction accuracy as criteria. Experimental results showed that the proposed model can produce the lowest prediction error and the highest prediction accuracy and outperformed the Wavelet-BPN, single BPN and random walk models. According to the experiments, it can be concluded that the proposed method can effectively detect and remove the noise from stock prices/indices and improve the forecasting performance of BPN. It is believed that the proposed method can be applied to other problem domains that tend to have high noise data. Future researches may aim at combining ICA and other forecasting tools, like support vector regression (SVR) and multivariate adaptive regression splines (MARS), in evaluating the ability of the proposed denoising scheme. Integrating BPN and other signal processing techniques, like non-negative matrix factorization (NMF), in further improving the forecasting capabilities can also be investigated in future studies.

Acknowledgements

This research was partially supported by the National Science Council of the Republic of China under Grant No. NSC 97-2221-E-231-008.

References

Assume, A., Campbell, J., & Murtage, F. (1998). Wavelet-based feature extraction and decomposition strategies for financial forecasting. *Journal of Computational Intelligence in Finance, 6*(2), 5–12.

Atsalakis, G. S., & Valavanis, K. P. (2009). Surveying stock market forecasting techniques – Part II: Soft computing methods. *Expert Systems with Applications, 36*(3), 5932–5941.

Back, A., & Weigend, A. (1997). Discovering structure in finance using independent component analysis. In *Proceeding of fifth international conference on neural networks in capital market* (pp. 15–17).

Balachandher, K. G., Fauzias, M. N., & Lai, M. M. (2002). An examination of the random walk model and technical trading rules in the Malaysian stock market. *Quarterly Journal of Business and Economics, 41*, 81–104.

Beckmann, C. F., & Smith, S. M. (2004). Probabilistic independent component analysis for functional magnetic resonance imaging. *IEEE Transactions on Medical Imaging, 23*, 137–152.

Bell, A. J., & Sejnowski, T. J. (1995). An information-maximization approach to blind separation and blind deconvolution. *Neural Computation, 7*, 1129–1159.

Box, G. E. P., & Jenkins, G. M. (1970). *Time aeries analysis, forecasting and control*. San Francisco: Holden-Day.

Cao, Q., & Parry, M. E. (2009). Neural network earnings per share forecasting models: A comparison of backward propagation and the genetic algorithm. *Decision Support Systems, 47*(1), 32–41.

Chang, P. C., Liu, C. H., Lin, J. L., Fan, C. Y., & Ng, C. S. P. (2009). A neural network with a case based dynamic window for stock trading prediction. *Expert Systems with Applications, 36*(3), 6889–6898.

Chang, S. G., Yu, B., & Vetterli, M. (2000). Spatially adaptive wavelet thresholding with context modeling for image denoising. *IEEE Transactions on Image Processing, 9*(9), 1522–1531.

Chauvin, Y., & Rumelhart, D. E. (1995). *Backpropagation: Theory, architectures, and applications*. New Jersey: Lawrence Erlbaum Associates.

Chavarnakul, T., & Enke, D. (2008). Intelligent technical analysis-based equivolume charting for stock trading using neural networks. *Expert Systems with Applications, 34*(2), 1004–1017.

Cheng, B., & Titterington, D. M. (1994). Neural network: A review from a statistical perspective (with discussion). *Statistical Science, 9*, 2–54.

Cheung, Y. M., & Xu, L. (2001). Independent component ordering in ICA time series analysis. *Neurocomputing, 41*, 145–152.

Cichocki, A., & Amari, S. I. (2002). *Adaptive blind signal and image processing: Learning algorithms and applications*. New York: John Wiley & Sons.

David, V., & Sanchez, A. (2002). Frontiers of research in BSS/ICA. *Neruocomputing, 49*, 7–23.

Déniz, O., Castrillón, M., & Hernández, M. (2003). Face recognition using independent component analysis and support vector machines. *Pattern Recognition Letters, 24*, 2153–2157.

Donoho, D. L., & Johnstone, I. M. (1998). Minimax estimation via wavelet shrinkage. *Annals of Statistics, 26*(3), 879–921.

Enke, D., & Thawornwong, S. (2005). The use of data mining and neural networks for forecasting stock market returns. *Expert Systems with Applications, 29*(4), 927–940.

Gonzalez, R. C., & Woods, R. E. (2002). *Digital image processing*. Englewood Cliffs, NJ: Prentice-Hall.

Górriz, J. M., Puntonet, C. G., Salmerón, M., & Lang, E. (2003). Time series prediction using ICA algorithms. *Computing, International Scientific Journal, 2*(2), 69–75.

Hall, J. W. (1994). Adaptive selection of US stocks with neural nets. In G. J. Deboeck (Ed.), *Trading on the edge: Neural, genetic and fuzzy systems for chaotic financial markets* (pp. 45–65). New York: Wiley.

Hassan, M. R., Nath, B., & Kirley, M. (2007). A fusion model of HMM, ANN and GA for stock market forecasting. *Expert Systems with Applications, 33*(1), 171–180.

Hussain, M. S., Reaz, M. B. I., Mohd-Yasin, F., & Ibrahimy, M. I. (2009). Electromyography signal analysis using wavelet transform and higher order statistics to determine muscle contraction. *Expert Systems, 26*, 35–48.

Haykin, S. (1999). *Neural network: A comprehensive foundation*. Englewood Cliffs, NJ: Prentice-Hall.

Hyvärinen, A., & Oja, E. (2000). Independent component analysis: Algorithms and applications. *Neural Networks, 13*, 411–430.

Hyvärinen, A., Karhunen, J., & Oja, E. (2001). *Independent component analysis*. New York: John Wiley and Sons.

James, C. J., & Gibson, O. J. (2003). Temporally constrained ICA: An application to artifact rejection in electromagnetic brain signal analysis. *IEEE Transactions on Biomedical Engineering, 50*, 1108–1116.

Jang, G. J., Lee, T. W., & Oh, Y. H. (2002). Learning statistically efficient features for speaker recognition. *Neurocomputing, 49*, 329–348.

Kim, K. J. (2006). Artificial neural networks with evolutionary instance selection for financial forecasting. *Expert Systems with Applications, 30*(3), 519–526.

Kim, T. K., Kim, H., Hwang, W., & Kittler, J. (2004). Independent component analysis in a local facial residue space for face recognition. *Pattern Recognition, 37*, 1873–1885.

Kiviluoto, K., & Oja, E. (1998). Independent component analysis for parallel financial time series. In *Proceeding of the fifth international conference on neural information* (pp. 895–898). Tokyo, Japan.

Lee, T. W. (1998). *Independent component analysis: Theory and application*. Boston: Kluwer Academic Publishers.

Lee, T. S., & Chen, N. J. (2002). Investigating the information content of non-cash-trading index futures using neural networks. *Expert Systems with Applications, 22*, 225–234.

Lee, T. S., & Chiu, C. C. (2002). Neural network forecasting of an opening cash price index. *International Journal of Systems Science, 33*, 229–237.

Leigh, W., Hightower, R., & Modani, N. (2005). Forecasting the New York stock exchange composite index with past price and interest rate on condition of volume spike. *Expert Systems with Applications, 28*(1), 1–8.

Li, T., Li, Q., Zhu, S., & Ogihara, M. (2003). A survey on wavelet applications in data mining. *SIGKDD Explorations, 4*(2), 49–68.

Malaroiu, S., Kiviluoto, K., & Oja, E. (2000), Time series prediction with independent component analysis. In *Proceedings of international conference on advanced investment technology* (pp. 895–898). Australia: Gold Coast.

McNelis, P. D. (2004). *Neural networks in finance: Gaining predictive edge in the market*. New York: Academic Press.

Oja, E., Kiviluoto, K., & Malaroiu, S., 2000. Independent component analysis for financial time series. In *Proceeding of the IEEE 2000 adaptive systems for signal processing, communications, and control symposium* (pp. 111–116). Lake Louise, Canada.

Popescu, T. D. (2003). Multivariate time series forecasting using independent component analysis. In *Proceedings of IEEE conference on emerging technologies and factory automation* (Vol. 2, pp. 782–789). Lisbon, Portugal.

Ramsey, J. B. (2002). Wavelets in economics and finance: Past and future. *Studies in Nonlinear Dynamics and Econometrics, 6*(3), 1–27.

Rao, R. M., & Bopardikar, A. S. (1998). *Wavelet transforms: Introduction to theory and applications*. Boston: Addison Wesley.

Stern, H. S. (1996). Neural networks in applied statistics. *Technometrics, 38*, 205–216.

Tsang, P. M., Kwok, P., Choy, S. O., Kwan, R., Ng, S. C., Mak, J., et al. (2007). Design and implementation of NN5 for Hong Kong stock price forecasting. *Engineering Applications of Artificial Intelligence, 20*(4), 453–461.

Vellido, A., Lisboa, P. J. G., & Vaughan, J. (1999). Neural networks in business: A survey of applications (1992–1998). *Expert Systems with Applications, 17*, 51–70.

Wood, S. (2002). *Float analysis: Powerful technical indicators using price and volume*. New York: John Wiley and Sons.

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

Yu, R. Q., Wang, J. H., & Jiang, J. H. (1996). Robust backpropagation algorithm as chemometric tool to prevent the over-fitting to outliers. *Chemometrics and Intelligent Laboratory Systems, 34*, 109–115.

Yudong, Z., & Lenan, W. (2009). Stock market prediction of S&P 500 via combination of improved BCO approach and BP neural network. *Expert Systems with Applications, 36*(5), 8849–8854.

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

Zhao, W., Chen, D., & Hu, S. (2004). Detection of outlier and a robust BP algorithm against outlier. *Computers and Chemical Engineering, 28*, 1403–1408.

Zheng, G., Starck, J. L., Campbell, J. G., & Murtagh, F. (1999). The wavelet transform for filtering financial data streams. *Journal of Computational Intelligence in Finance, 7*, 18–35.

Zhu, X., Wang, H., Xu, L., & Li, H. (2008). Predicting stock index increments by neural networks: The role of trading volume under different horizons. *Expert Systems with Applications, 34*(4), 3043–3054.