Financial time series forecasting using independent component analysis and support vector regression

Chi-Jie Lu [a], Tian-Shyug Lee [b,*], Chih-Chou Chiu [c]

[a] Department of Industrial Engineering and Management, Ching Yun University, Taiwan, ROC
[b] Graduate Institute of Management, Fu Jen Catholic University, Taiwan, ROC
[c] Institute of Commerce Automation and Management, National Taipei University of Technology, Taiwan, ROC

ARTICLE INFO

ABSTRACT

As financial time series are inherently noisy and non-stationary, it is regarded as one of the most challenging applications of time series forecasting. Due to the advantages of generalization capability in obtaining a unique solution, support vector regression (SVR) has also been successfully applied in financial time series forecasting. In the modeling of financial time series using SVR, one of the key problems is the inherent high noise. Thus, detecting and removing the noise are important but difficult tasks when building an SVR forecasting model. To alleviate the influence of noise, a two-stage modeling approach using independent component analysis (ICA) and support vector regression is proposed in financial time series forecasting. ICA is a novel statistical signal processing technique that was originally proposed to find the latent source signals from observed mixture signals without having any prior knowledge of the mixing mechanism. The proposed approach first uses ICA to the forecasting variables for generating 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 which contain less noise and served as the input variables of the SVR forecasting model. In order to evaluate the performance of the proposed approach, the Nikkei 225 opening index and TAIEX closing index are used as illustrative examples. Experimental results show that the proposed model outperforms the SVR model with non-filtered forecasting variables and a random walk model.

© 2009 Elsevier B.V. All rights reserved.

1. Introduction

There has been growing interest in financial time series forecasting in recent years as accurate forecasting of financial prices/indices has become an important issue in investment decision making. However, financial time series are inherently noisy and non-stationary [19,64]. The noise characteristic refers to the unavailability of complete information from past behavior of financial markets to fully capture the dependency between future and past prices. The information that is not included in the forecasting model is considered as noise while the non-stationary characteristic implies that the distribution of financial time series is changing over time. Therefore, financial time series forecasting is regarded as one of the most challenging tasks of time series forecasting.

Neural networks have been found to be useful techniques for modeling financial time series due to their ability to capture subtle functional relationships among the empirical data even though the underlying relationships are unknown or hard to describe [34,36–38,52,61,65,66]. Unlike traditional statistical models, known as Box-Jenkins ARIMA [5], neural networks 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 [20,61,66]. The most popular neural network training algorithm for financial forecasting is the backpropagation neural networks (BPN) that has a simple architecture but a powerful problem-solving ability. However, the BPN also suffers from a number of shortcomings such as the need for a large number of controlling parameters, difficulty in obtaining a stable solution and the risk of model over-fitting [7,8,55,56].

Support vector machines (SVMs) is a novel neural network algorithm based on statistical learning theory [59,60]. It can lead to great potential and superior performance in practical applications. This is largely due to the structure risk minimization principles in SVMs, which has greater generalization ability and is superior to the empirical risk minimization principle as adopted by traditional neural networks. Due to the advantages of the generalization capability in obtaining a unique solution, the SVMs have drawn the attention of researchers and have been applied in many applications such as texture classification, image recognition, data mining and bioinformatics [6,14,22,31,40,44,46,50]. With the introduction of Vapnik's ε-insensitivity loss function, the regression model of SVMs, called support vector regression (SVR), has also been receiving increasing attention to solve nonlinear estimation problems [59,60]. It has been successfully applied in different problems of time series prediction

* Corresponding author. Tel.: +886 2 2905 2905; fax: +886 2 2901 8475.
E-mail addresses: 036665@mail.fju.edu.tw, mleetw@yahoo.com (T.-S. Lee).

such as production value forecast of machinery industry, engine reliability prediction, wind speed prediction and financial time series forecasting [7,8,26,30,45,48,55–57]. Since there are many successful results of utilizing SVR in time series prediction, it motivates our research work by using SVR for financial time series forecasting.

In the modeling of financial time series using SVR, one of the key problems is the inherent noise of the financial time series. 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 [7,19]. Therefore, detecting and removing the noise are important but difficult tasks when building an SVR forecasting model. Few studies have been proposed to deflate the influence of noisy data and enhance the robust capability of SVR. Chuang et al. [15] proposed a robust support vector regression network. They used the concept of tradition robust statistics to fine tune the model obtained by SVR trying to reduce the overfitting phenomenon and improve the learning performance. Suykens et al. [53] presented a weighted version of least squares SVM (LS-SVM) to overcome the effects of outliers. In their approach, an LS-SVM was trained on the entire dataset for yielding the support values. A small fraction of the dataset associated with support values having the smallest magnitude are discarded and the LS-SVM retrained on the remaining data. This process is repeated until a sufficiently small kernel expansion is obtained.

As the existing methods would either involve extensive computation or use additional parameters in SVR algorithm to reduce the effects of outliers/noise contained in the data. However, the consuming time of performing SVR algorithm will be increased while the extensive computation is carried out. 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 avoid the limitations of the existing method and reduce the influence of noise, a two-stage approach by combining independent component analysis (ICA) and support vector regression is proposed in this research for modeling financial time series.

ICA is a novel statistical signal processing technique to find independent sources given only observed data that are mixtures of unknown sources without any prior knowledge of the mixing mechanism [25,35]. In the basic ICA model, the observed mixture signals $\mathbf{X}$ can be expressed as $\mathbf{X} = \mathbf{AS}$, where $\mathbf{A}$ is an unknown mixing matrix and $\mathbf{S}$ represents the latent source signals that cannot be directly observed from the mixture signals $\mathbf{X}$. The ICA model describes how the observed mixture signals are generated by a process that uses the mixing matrix $\mathbf{A}$ to linearly mix the latent source signals $\mathbf{S}$. The source signals are assumed to be mutually statistically independent. Based on this assumption, the ICA solution is obtained in an unsupervised learning process that finds a de-mixing matrix $\mathbf{W}$. The de-mixing matrix $\mathbf{W}$ is used to transform the observed mixture signals $\mathbf{X}$ to yield the independent signals $\mathbf{Y}$, i.e., $\mathbf{WX} = \mathbf{Y}$. The independent signals $\mathbf{Y}$ are then used as the estimates of the latent source signals $\mathbf{S}$. The rows of $\mathbf{Y}$, called independent components (ICs), are required to be as mutually independent as possible. Even though the basic ICA model has been widely applied in signal processing, face recognition, feature extraction and quality control [3,17,28,29,32,43,42,58], there are still few applications using ICA in financial time series forecasting.

Back and Weigend [1] 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 [33] employed ICA to find the fundamental factors affecting the cash flow of 40 stores in the same retail chain. They found that the cash flow of the retail stores was mainly affected by holidays, seasons and competitors' strategies. Oja et al. [47] applied ICA in foreign exchange rate time forecasting. They first 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 linear and nonlinear smoothing techniques. Then, the autoregression (AR) model 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 predictions for the original observed time series.

There are only very few articles addressing both ICA and SVR in conducting forecasting tasks. Cao and Chong [9] employed ICA as a feature extraction tool in developing a SVM forecaster. The independent components (ICs) were considered as features of the forecasting data and used to build the SVM forecasting model. Chen et al. [11] combined dynamic independent component analysis (DICA) with SVR to construct multi-layer support vector regression model. The DICA was used in the first layer to extract the major dynamic features from the process. The second layer is the SVR that makes the regression estimation based the extracted features. Hou et al. [21] applied ICA and SVR in near-infrared (NIR) spectral analysis. They used ICA to extract the independent components and corresponding mixing matrix from the NIR spectra of chemical components, then the SVR was used to build a model between mixing matrix and the real concentration matrix of chemical components for spectral analysis. Wang et al. [62] utilized kernel independent component analysis and SVR for the estimation of source ultraviolet spectra profiles and simultaneous determination of polycomponents in mixtures. They applied ICA to estimate the ultraviolet source spectra profiles. Then, the calibration model was build by using SVR based on the mixing matrix. The existing ICA–SVR model approach usually only uses independent components or the mixing matrix as the inputs of the built SVR model. Moreover, the existing method did not discuss the features of the ICs. On the other hand, our proposed ICA–SVR model identifies the ICs that can be used to represent the main feature or noise of the original data. Based on these two points, we believe that our proposed modeling approach differs from those appeared in the literature and hence provides an ideal alternative in conducting financial time series forecasting.

In this study, we present a financial time series forecasting model by integrating ICA and SVR. The ICA method is used to detect and remove the noise of financial time series data and further improve the performance of SVR. The proposed approach first uses ICA to the forecasting variables to estimate the independent components and mixing matrix. Since the financial time series 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 which contain less noise. The SVR then uses the filtered (or de-noised) forecasting variables to build the forecasting model. In order to evaluate the performance of the proposed approach, the Nikkei 225 opening cash index and TAIEX (Taiwan Stock Exchange Capitalization Weighted Stock Index) closing cash index are used as the illustrative examples.

The rest of this paper is organized as follows. Sections 2 and 3 respectively, give a brief introduction about independent component analysis and support vector regression. The proposed two-stage forecasting model is thoroughly described in Section 4. Section 5 presents the experimental results from the datasets including the Nikkei 225 opening cash index and TAIEX closing cash index. The paper is concluded in Section 6.

2. Independent component analysis

Let $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_m]^T$ be a multivariate data matrix of size $m \times n$, $m \leq n$, consisting of observed mixture signals $\mathbf{x}_i$ of size $1 \times n$, $i = 1, 2, ..., m$. In the basic ICA model, the matrix $\mathbf{X}$ can be modeled as [24]

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

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 observed mixture signals $\mathbf{x}_i$. The ICA model aims at finding an $m \times m$ de-mixing matrix $\mathbf{W}$ such that

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

where $\mathbf{y}_i$ is the ith row of the matrix $\mathbf{Y}$, $i = 1,2,...,m$. For using vectors $\mathbf{y}_i$ to estimate the independent latent source signals ($\mathbf{s}_i$), $\mathbf{y}_i$ must be statistically independent, 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 [4,16,23]. In general, the ICs are obtained by using the de-mixing matrix $\mathbf{W}$ to multiply the matrix $\mathbf{X}$, i.e. $\mathbf{Y} = \mathbf{W}\mathbf{X}$. 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 [25], and the non-Gaussianity of the ICs can be measured by the negentropy:

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

where $\mathbf{y}_{\text{gauss}}$ 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 $H(\mathbf{y}) = -\int p(\mathbf{y})\log p(\mathbf{y})\mathrm{d}\mathbf{y}$.

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 [24]:

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

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 nonquadratic function, and is given by $G(y) = \exp(-y^2/2)$ in this study. The *FastICA* algorithm proposed by [23] is adopted in this paper to solve for the de-mixing matrix $\mathbf{W}$.

Two preprocessing steps are common in ICA modeling, centering and whitening [24]. First, the input matrix $\mathbf{x}$ is centered by subtracting the row means of the input matrix, i.e., $\mathbf{x}_i \leftarrow (\mathbf{x}_i - E(\mathbf{x}_i))$. The matrix $\mathbf{X}$ with zero mean is then passed through the whitening matrix $\mathbf{V}$ to remove the second order statistic of the input matrix, i.e., $\mathbf{Z} = \mathbf{V}\mathbf{X}$. The whitening matrix $\mathbf{V}$ is twice the inverse square root of the covariance matrix of the input matrix, i.e., $\mathbf{V} = 2 \cdot (C_x))^{-(1/2)}$, where $C_x = E(\mathbf{x}\mathbf{x}^T)$ is the covariance matrix of $\mathbf{X}$. The rows of the whitened input matrix $\mathbf{Z}$, denoted by $\mathbf{z}$, are uncorrelated and have unit variance, i.e., $E(\mathbf{z}\mathbf{z}^T) = \mathbf{I}$. In this study, it is assumed that the training and testing financial time series datasets are centered and whitened.

3. Support vector regression

The support vector regression is an adaptation of recently introduced statistical/ machine learning theory based classification paradigm namely, support vector machines. For illustrating the concept of SVR, a typical regression problem is formulated. Consider a set of data $G = \{(\mathbf{x}_i, q_i)\}_{i=1}^n$, where $\mathbf{x}_i$ is a vector of the model inputs, q_i is actual value and represents the corresponding scalar output, and n is total number of data patterns. The objective of the regression analysis is to determine a function $f(\mathbf{x})$, so as to predict accurately the desired (target) outputs (q). Thus, the typical regression function can be formulated as $q_i = f(\mathbf{x}_i) + \delta$, where δ is the random error with distribution of $N(0, \sigma^2)$. The regression problem can be classified as linear and nonlinear regression problems. As the nonlinear regression problem is more difficult to deal with, SVR was mainly developed for tackling the nonlinear regression problem.

To solve a nonlinear regression problem, in SVR, the inputs are first nonlinearly mapped into a high dimensional feature space ($\mathbf{F}$) wherein they are correlated linearly with the outputs. The SVR formalism considers the following linear estimation function [59,60]:

$$f(\mathbf{x}) = (\mathbf{v} \cdot \Phi(\mathbf{x})) + b \tag{5}$$

where, $\mathbf{v}$ is weight vector, b is a constant, $\Phi(\mathbf{x})$ denotes a mapping function in the feature space, and $(\mathbf{v} \cdot \Phi(\mathbf{x}))$ describes the dot production in the feature space $\mathbf{F}$. In SVR, the problem of nonlinear regression in the lower dimension input space ($\mathbf{x}$) is transformed into a linear regression problem in a high dimension feature space ($\mathbf{F}$). That is, the original optimization problem involving a nonlinear regression is recast as searching the flattest function in the feature space, and not in the input space.

A number of cost functions such as the Laplacian, Huber's Gaussian and ε-insensitive can be used in the SVR formulation. Among these, the robust ε-insensitive loss function (L_ε), given below is the most commonly adopted [60].

$$L_\varepsilon(f(\mathbf{x}), q) = \begin{cases} |f(\mathbf{x}) - q| - \varepsilon & \text{if } |f(\mathbf{x}) - q| \geq \varepsilon \\ 0 & \text{otherwise} \end{cases} \tag{6}$$

where ε is a precision parameter representing the radius of the tube located around the regression function $f(\mathbf{x})$ (see the broken lines in Fig. 1). A schematic representation of the SVR using ε-insensitive loss function is illustrated in Fig. 1. In Fig. 1, the region enclosed by the tube is known as "ε-insensitive zone" since the loss function assumes a zero value in this region and does not penalize the prediction errors with magnitudes smaller than ε.

The weight vector ($\mathbf{v}$) and constant (b) in Eq. (5) can be estimated by minimizing the following regularized risk function [60]:

$$R(C) = C\frac{1}{n}\sum_{i=1}^n L_e(f(\mathbf{x}_i), q_i) + \frac{1}{2}|\mathbf{w}|^2 \tag{7}$$

where $L_e(f(\mathbf{x}), q)$ is ε-insensitive loss function in Eq. (6); $\frac{1}{2}|\mathbf{w}|^2$ is the regularization term which controls the trade-off between the complexity and the approximation accuracy of the regression model to ensure that the model possesses an improved generalized performance; C is the regularization constant used to specify the trade-off between the empirical risk and regularization term. Both C and ε are user-determined parameters.
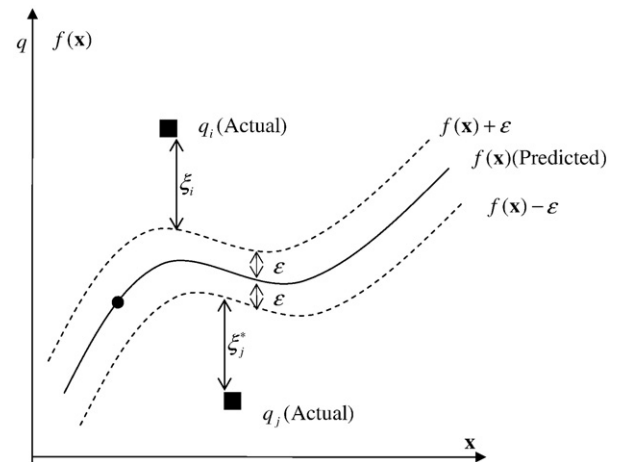

Fig. 1. A schematic representation of the SVR using ε-insensitive loss function.
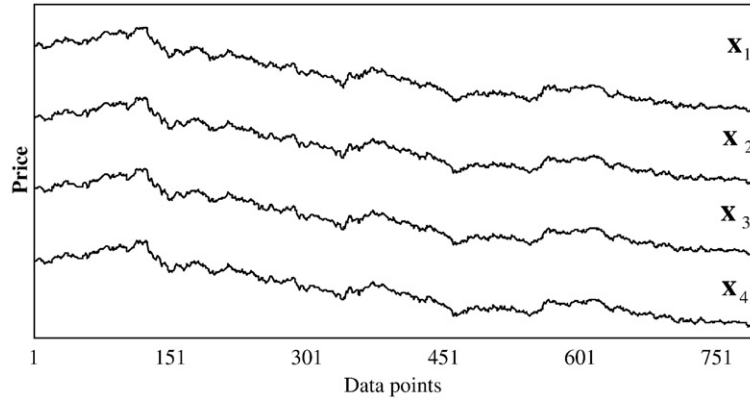
Fig. 2. Four financial time series data of size 1×794, respectively.

Two positive slack variables, ξ_i and ξ_i^*, $i = 1,2,...,n$, can be used to measure the deviation $(q_i - f(\mathbf{x}_i))$ from the boundaries of the ε-insensitive zone. That is, they represent the distance from actual values to the corresponding boundary values of ε-insensitive zone (see Fig. 1). By using slack variables, the Eq. (7) is transformed into the following constrained form:

$$\text{Minimize:} \quad R_{\text{reg}}(f) = \frac{1}{2} \mid \mathbf{w} \mid^2 + C \sum_{i=1}^{n} \left(\xi_i + \xi_i^* \right) \tag{8}$$

subject to,

$$\begin{cases} q_i - (\mathbf{w} \cdot \boldsymbol{\Phi}(\mathbf{x}_i)) - b \leq \varepsilon + \xi_i \\ (\mathbf{w} \cdot \boldsymbol{\Phi}(\mathbf{x}_i)) + b - q_i \leq \varepsilon + \xi_i^* \\ \xi_i, \xi_i^* \geq 0, \quad \text{for } i = 1, ..., n \end{cases}$$

By using Lagrangian multipliers and Karush–Kuhn–Tucker conditions to the Eq. (8), it thus yields the following dual Lagrangian form [60],

Maximize:

$$\begin{aligned} L_{\text{d}}\left(\alpha, \alpha^*\right) = &- \varepsilon \sum_{i=1}^{n} \left(\alpha_i^* + \alpha_i\right) + \sum_{i=1}^{n} \left(\alpha_i^* - \alpha_i\right) q_i \\ &- \frac{1}{2} \sum_{i,j=1}^{n} \left(\alpha_i^* - \alpha_i\right)\left(\alpha_j^* - \alpha_j\right) \mathbf{K}\left(\mathbf{x}_i, \mathbf{x}_j\right) \end{aligned} \tag{9}$$

subject to the constraints,

$$\begin{cases} \sum_{i=1}^{n} \left(\alpha_i^* - \alpha_i\right) = 0 \\ 0 \leq \alpha_i \leq C, \ i = 1, ..., n \\ 0 \leq \alpha_i^* \leq C, \ i = 1, ..., n \end{cases}$$

The Lagrangian multipliers in Eq. (9) satisfy the equality $\alpha_i \alpha_i^* = 0$. The Lagrangian multipliers, α_i and α_i^*, are calculated and an optimal desired weight vector of the regression hyperplan is, $\mathbf{v}^* = \sum_{i=1}^{n} \left(\alpha_i - \alpha_i^*\right) K(\mathbf{x}, \mathbf{x}_i)$. Hence, the general form of the SVR-based regression function can be written as [60],

$$f(\mathbf{x}, \mathbf{v}) = f\left(\mathbf{x}, \alpha, \alpha^*\right) = \sum_{i=1}^{n} \left(\alpha_i - \alpha_i^*\right) K(\mathbf{x}, \mathbf{x}_i) + b \tag{10}$$

where $K(\mathbf{x}, \mathbf{x}_i)$ is called the kernel function. The values of the kernel equals the inner product of two vectors, $\mathbf{x}_i$ and $\mathbf{x}_j$, in the feature space $\Phi(\mathbf{x}_i)$ and $\Phi(\mathbf{x}_j)$; that is, $K(\mathbf{x}_i, \mathbf{x}_j) = \Phi(\mathbf{x}_i)\Phi(\mathbf{x}_j)$. Any function that meets Mercer's condition can be used as the kernel function [59,60]. Although several choices for the kernel function are available, the most widely used kernel unction is the radial basis function (RBF) defined as $K(\mathbf{x}_i, \mathbf{x}_j) = exp\left(\frac{-||\mathbf{x}_i - \mathbf{x}_j||^2}{2\sigma^2}\right)$ [12,60], where σ denotes the width of the RBF. Thus, the RBF is applied in this study as kernel function.

4. Proposed forecasting model using ICA and SVR

For the proposed two-stage forecasting method, ICA is first applied to filter out the noise contained in forecasting variables. The filtered forecasting variables are then used in SVR for constructing a forecasting model. When using ICA for de-noising, the basic ICA model is first utilized to the mixture matrix **X** of size $m \times n$ combined from m forecasting variables ($\mathbf{x}_i$) of size $1 \times n$ for estimating a de-mixing matrix (**W**) of size $m \times m$ and independent components ($\mathbf{y}_i$) of size $1 \times n$. Since the noise of time series data may contain least information about the trend of time series, the ICs which cannot capture the trend of time series 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 [13], using *Relative*
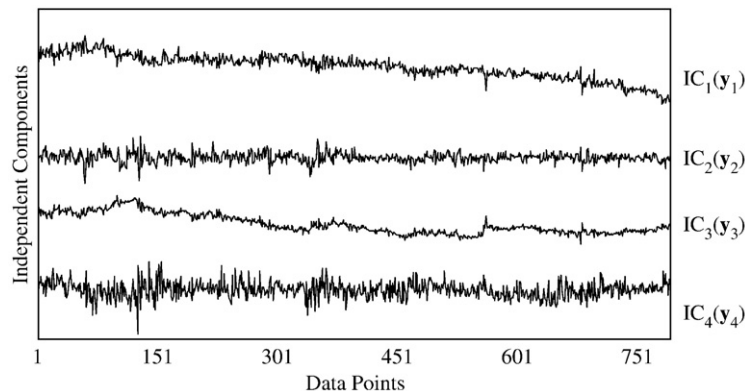

Fig. 3. Four ICs of the time series data in Fig. 2.

Table 1
The *RHD* reconstruction errors of every iteration of *TnA* approach for ordering the ICs in Fig. 3.

Iterations	The ICs included in the matrix ($\hat{X}$)	*RHD* reconstruction error
1	IC$_1$,IC$_2$,IC$_3$	0.7662
	IC$_1$,IC$_2$,IC$_4$	1.7181
	IC$_1$,IC$_3$,IC$_4$	1.4294
	IC$_2$,IC$_3$,IC$_4$	1.8948
2	IC$_1$,IC$_2$	1.7137
	IC$_1$,IC$_3$	1.3346
	IC$_2$,IC$_3$	1.9158
3	IC$_1$	1.7531
	IC$_3$	1.9779

Hamming Distance (*RHD*) reconstruction error as index is adapted to order the ICs. The *RHD* reconstruction error can evaluate the similarity between time series data. The smaller *RHD* value represents the higher similarity between time series data. The *RHD* value of identical time series data is zero, whereas the *RHD* value of totally different time series data is four. The equation of *RHD* is given in Appendix A.

An example is used for illustrating the concept of the *TnA* method. Fig. 2 shows four financial time series data, each of size 1×794, which can be combined as a mixture matrix **X** of size 4×794. After using ICA method to the matrix **X**, a de-mixing matrix **W** of size 4×4 and four ICs, each of size 1×794, can be estimated. The profiles of those four 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 four ICs on capturing the main feature of the data in Fig. 2, the *TnA* approach is then used.

Consider a set of m ICs. 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 one IC in the ordering, the reconstructed mixture matrix $\hat{\mathbf{X}}$ can be obtained by using the following equation,

$$\hat{X} = \sum_{i=1, i \neq k}^{m} \mathbf{a}_i \mathbf{y}_i, \ 1 \le k \le m \tag{11}$$

where $\hat{\mathbf{X}} = [\hat{\mathbf{x}}_1, \hat{\mathbf{x}}_2, \ldots, \hat{\mathbf{x}}_m]^T$ is reconstructed mixture matrix of size $m \times n$, $\hat{\mathbf{x}}_i$ is reconstructed forecasting variable, $\mathbf{a}_i$ is the i-th column vector of mixing matrix **A**, $\mathbf{A} = \mathbf{W}^{-1}$, and $\mathbf{y}_i$ is the i-th IC. After obtaining the reconstructed matrices considering different IC as the last one IC in the ordering, the *RHD* reconstruction error between each reconstructed matrix $\hat{\mathbf{X}}$ and original mixture matrix **X** can be computed. More precisely, the *RHD* value is used to evaluate the similarity between the original forecasting variables ($\mathbf{x}_i$) and corresponding reconstructed forecasting variables ($\hat{\mathbf{x}}_i$). For example, If the RHD value between the variable $\mathbf{x}_1$ and its specific reconstructed variable $\hat{\mathbf{x}}_1$ is close to zero,

the main feature of these two variables are similar. The ICs used to reconstruct the reconstructed variable $\hat{\mathbf{x}}_1$ contain the main feature of the variable $\mathbf{x}_1$. Conversely, the *RHD* value between the variable $\mathbf{x}_1$ and the variable $\hat{\mathbf{x}}_1$ is far from zero. The corresponding utilized ICs are considered as including less main information of the variable $\mathbf{x}_1$.

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 $\hat{\mathbf{X}}$ excluded IC$_4$ has the smallest *RHD* value in the first iteration. This indicates that IC$_4$ contributes the least information in the reconstruction of original data and contains less information about the main feature of the original forecasting variables than that of the remaining ICs. Thus, IC$_4$ is selected as the last IC in the ordering and removed from the sorting process before the next iteration. In the next iterations, the *TnA* algorithm repeat the same operations using Eq. (11) on the remaining $m - 1$ independent components, *i.e.* 3 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 IC$_2$ since it has the smallest *RHD* value in the second iteration. According to Table 1, the four ICs can be ordered as follows: IC$_1$, IC$_3$, IC$_2$ and IC$_4$. The first one IC, *i.e.* IC$_1$, contains 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.* IC$_4$, 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 IC$_1$, IC$_2$ and IC$_3$ for reconstruction. Fig. 4 shows the first 201 data points of the series data in Fig. 2 and its two reconstructed series data respectively using IC$_1$, IC$_2$ and IC$_3$, and only IC$_4$. It can be seen from Fig. 4 that the trend and shape of the reconstructed series data using IC$_1$, IC$_2$ and IC$_3$ are very similar to its original series data, i.e. $\mathbf{x}_1$. Conversely, the reconstructed series data using only IC$_4$ is very different from the original series data.

After obtaining the de-noised forecasting variables, they are then used in building the SVR financial time series forecasting model. The first step of using SVR is the selection of kernel function. As mentioned in Section 3, the RBF kernel function is adapted in this study. It is well known that the performance (estimation accuracy) of SVR depends on setting of parameters. Thus, the selection of three parameters, regularization constant C, loss function ε and σ (the width of the RBF) of a SVR model is important to the accuracy of forecasting.

The performance of SVR is mainly affected by the setting of parameters C and ε [12,41]. There are no general rules for the choice of C and ε. The selection is usually based on trial-and-error (or called
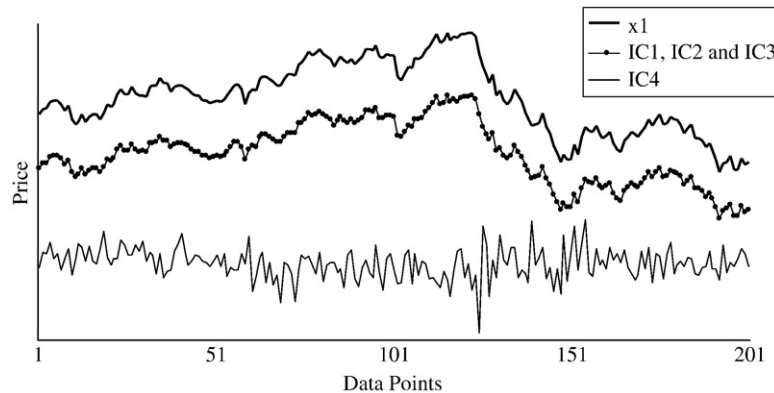
Fig. 4. The first 201 data points of the series data (thick line) in Fig. 2 and its two reconstructed series data respectively using IC$_1$, IC$_2$ and IC$_3$ (dotted line) and only IC$_4$ (thin line).
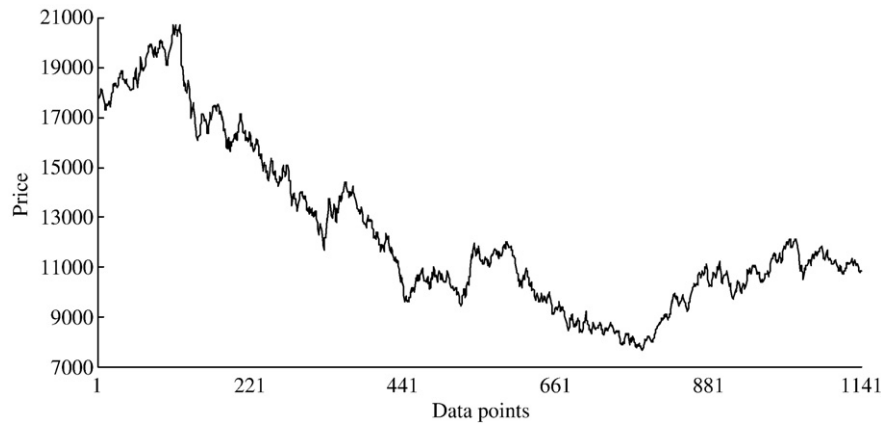
Fig. 5. The daily Nikkei 225 opening cash prices from October 4, 1999 to September 30, 2004.

cross-validation) method or user's prior knowledge and/or expertise. However, the cross-validation method is very time consuming and data-intensive. On the other hand, the user's expertise method faces the risk of using single parameter set and is not appropriate for non-expert users. To alleviate the possible drawbacks mentioned above, an analytic parameter selection method proposed by Cherkassky and Ma [14] and the grid search proposed Lin et al. [41] are used in this study for parameters setting of C and ε. The analytic method is based on sketching the structure of training data to determine the best value of parameters. The grid search is a straightforward method using exponentially growing sequences of C and ε to identify good parameters (for example, $C = 2^{-5}, 2^{-3}, 2^{-1}, \ldots, 2^{15}$). In order to combine the advantages of those two approaches, we first use the analytic method to select a parameter set of C and ε. Then, the grid search uses the set as starting point for searching. The parameter set of C and ε which generate the minimum forecasting mean square error is considered as the best parameter set. Cherkassky and Ma [12] pointed out that for multivariate d-dimensional problems, the RBF width parameter σ is set as $\sigma^d \sim (0.1, 0.5)$, where d is number of input variables. For simplifying the setting of parameter, in this study, $\sigma = 0.8$ is used for all experiments. For the details of the analytic method, please refer to Cherkassky and Ma [12].

5. Empirical study

5.1. Datasets and performance criteria

For evaluating the performance of the proposed forecasting model using ICA and SVR (called ICA–SVR model), the daily Nikkei 225 opening cash index and TAIEX closing cash index are used in this study. 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 [36,37]. 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 SGX-DT (Singapore Exchange-Derivative Trading Limited), OSE (Osaka Securities Exchange) and CME (Chicago Mercantile Exchange) markets. The previous day's cash market closing index is also an important variable for predicting the cash market opening price [36,37]. Therefore, four forecasting variables are used for predicting the Nikkei 225 opening cash index. The daily data of futures and cash prices from October 4, 1999 to September 30, 2004 of the Nikkei 225 cash index provided by Bloomberg are collected in this study. There are totally 1144 data points in the dataset and the daily Nikkei 225 opening cash prices are shown in Fig. 5. The first 794 data points (69.41% of the total sample points) are used as the training sample while the remaining 350 data points (30.59% of the total sample points) are used as the testing sample.

For forecasting the TAIEX closing cash index, the TAIEX index futures prices and technical indicators are used as forecasting variables since technical indicators are the most widely used features in financial time series prediction [2,39]. There are two TAIEX index future contracts traded on SGX-DT and TAIFEX (Taiwan Futures Exchange) markets. The seven technical indicators, determined by the review of domain experts and literatures [39,63], are selected as forecasting variables for predicting the TAIEX closing cash index. The selected 7 technical indicators are the previous day's cash market
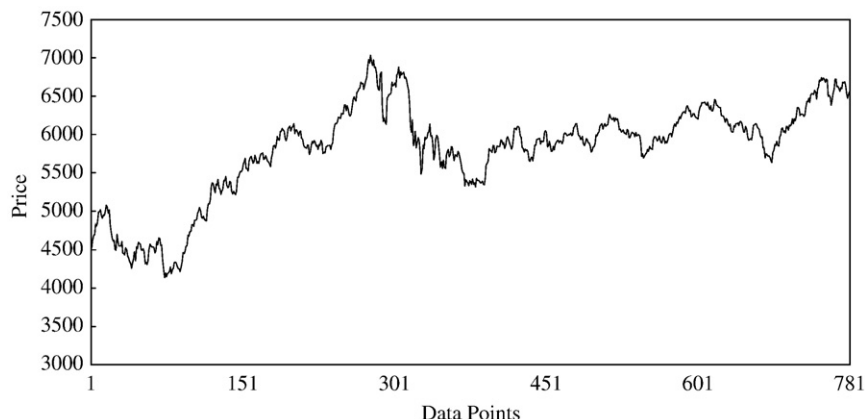

Fig. 6. The daily TAIEX closing cash prices from January 2003 to February 2006.

Table 2
Performance measures and their definitions.

Metrics	Calculation
RMSE	$RMSE = \sqrt{\dfrac{\sum_{i=1}^{n}(T_i - P_i)^2}{n}}$
NMSE	$NMSE = 1/(\sigma^2 n) * \sum_{i=1}^{n}(T_i - P_i)^2$, where $\sigma^2 = 1/(n-1) * \sum_{i=1}^{N}(P_i - \overline{P})^2$
MAD	$MAD = \dfrac{\sum_{i=1}^{n}|T_i - P_i|}{n}$
DS	$DS = \frac{100}{n}\sum_{i=1}^{n} d_i$, where $d_i = \begin{cases} 1 & (P_i - P_{i-1})(T_i - T_{i-1}) \geq 0 \\ 0 & \text{otherwise} \end{cases}$
CP	$CP = \frac{100}{n_2}\sum_{i=1}^{n} d_i$, where $d_i = \begin{cases} 1 & (T_i - T_{i-1}) > 0 \text{ and } (P_i - P_{i-1})(T_i - T_{i-1}) \geq 0 \\ 0 & \text{otherwise} \end{cases}$
CD	$CD = \frac{100}{n_2}\sum_{i=1}^{n} d_i$, where $d_i = \begin{cases} 1 & (T_i - T_{i-1}) < 0 \text{ and } (P_i - P_{i-1})(T_i - T_{i-1}) \geq 0 \\ 0 & \text{otherwise} \end{cases}$

Note that T and P represent the actual and predicted value, respectively, n is total number of data points, n_1 is number of data points belong to up trend and n_2 is number of data points belong to down trend.

high, low, amount and volume, and 6-days relative strength indicator (RSI 6), 10-days total amount weight stock price index (TAPI 10), and today's opening cash index. For the details about technical indicators, please refer to Balachandher et al. [2], Wood [63] and Leigh et al. [39]. Thus, 9 forecasting variables are used for TAIEX closing cash index forecasting. The daily data of futures, technical indicators, and cash prices from January 2, 2003 to February 27, 2006 of the TAIEX cash index provided by Capital Futures Corporation, Taipei, are collected as a dataset. The daily TAIEX closing cash prices in the TAIEX dataset are depicted in Fig. 6. There are totally 781 data points in the dataset. The first 546 data points (69.90% of the total sample points) are used as the training sample and the remaining 235 data points (30.10% of the total sample points) are used as testing sample.

The prediction performance is evaluated using the following performance measures, namely, the root mean square error (RMSE), normalized mean square error (NMSE), mean absolute difference (MAD), directional Symmetry (DS), correct up trend (CP) and correct down trend (CD). The definitions of these criteria can be found in Table 2. RMSE, NMSE and MAD are measures of the deviation between actual and predicted values. The smaller values of RMSE, NMSE and MAD, the closer are the predicted time series values to that of the actual value. They can be used to evaluate the prediction error. DS provides the correctness of the predicted direction of the cash index in terms of percentage. CP and CD provide the correctness of the predicted up trend and predicted down trend of the cash index in terms of percentage. DS, CP and CD can be utilized to evaluate the prediction accuracy.

Table 3
The model selection results of SVR model.

ε	C	Training MSE	Testing MSE
2^{-13}	2^{-1}	0.0000319491	0.0000248156
	2^1	0.0000291671	0.0000214102
	2^3	0.0000279661	0.0000240622
2^{-11}	2^{-1}	0.0000319861	0.0000247999
	2^1	**0.0000291736**	**0.0000214540**
	2^3	0.0000280469	0.0000239275
2^{-9}	2^{-1}	0.0000323087	0.0000273171
	2^1	0.0000294424	0.0000230685
	2^3	0.0000282242	0.0000221947
2^{-7}	2^{-1}	0.0000426816	0.0000335855
	2^1	0.0000346765	0.0000342352
	2^3	0.000031965	0.0000290501

Table 4
The model selection results of the proposed ICA–SVR model.

ε	C	Training MSE	Testing MSE
2^{-11}	2^{-1}	0.0000547405	0.0000279631
	2^1	0.0000535331	0.0000304249
	2^3	0.0000528961	0.0000396468
	2^5	0.0000525349	0.0000469127
2^{-9}	2^{-1}	0.0000546991	0.0000258684
	2^1	0.0000537528	0.0000284480
	2^3	0.0000529974	0.0000344395
	2^5	0.0000526340	0.0000400822
2^{-7}	2^{-1}	0.0000554437	0.0000202124
	2^1	0.0000544119	0.0000198950
	2^3	**0.0000538615**	**0.0000191902**
	2^5	0.0000533897	0.0000202790
2^5	2^{-1}	0.0001088290	0.0001002030
	2^1	0.0001415800	0.0001092080
	2^3	0.0001368980	0.0000906426
	2^5	0.0001514900	0.0001442630

5.2. Forecasting results of Nikkei 225 and TAIEX cash prices

The forecasting results of the proposed ICA–SVR model are compared to the SVR model using non-filtered forecasting variables (called single SVR model) and the random work model simply uses the previous day's price to predict today's price. For building SVR forecasting model, the *LIBSVM* package proposed by Chang and Lin [10] is adapted in this study. The original datasets are first scaled into the range of [−1.0, 1.0] when using LIBSVM package. The purpose of doing so is to ensure that large value input variables do not overwhelm smaller value inputs, thus helping to reduce prediction errors.

In the selection of parameters for modeling SVR, $C = 1.25$ and $\varepsilon = 0.0019$ can be obtained by the analytic approach mentioned in Section 4. Since $C = 1.25$ is near $C = 2^1$ and $\varepsilon = 0.0019$ is close to $\varepsilon = 2^{-9}$, the parameter set $(C = 2^1, \varepsilon = 2^{-9})$ is used as the starting point of grid search for searching the best parameters. The testing results of the SVR model with combinations of different parameter sets are summarized in Table 3. From Table 3, it can be found that the parameter set $(C = 2^1, \varepsilon = 2^{-11})$ gives the best forecasting result (minimum testing MSE) and is the best parameter set for SVR model in forecasting Nikkei 225 opening cash index.

In the modeling of the proposed ICA–SVR model, the noise of four forecasting variables should be removed first using ICA approach. As the four forecasting variables are the time series discussed and expressed in Fig. 2, the noise removing process and results have been discussed in Section 4. After using ICA to filter out the noise of the four forecasting variables, the de-noised forecasting variables are then used for building the SVR forecasting model. Using the same process when building the SVR model, the parameter set $(C = 2^1, \varepsilon = 2^{-9})$ obtained by the analytic method is used as the starting point of grid search. Table 4 summarizes the testing results of the proposed ICA–SVR model with combinations of different parameter sets. It can be observed from Table 4 that the parameter set $(C = 2^3, \varepsilon = 2^{-7})$ gives the best forecasting result and hence is the best parameter setup for the proposed ICA–SVR model in forecasting Nikkei 225 opening cash index.

The Nikkei 225 opening cash price index forecast results using random walk, SVR and the proposed ICA–SVR models are computed and listed in Table 5. From Table 5, it can be found that the RMSE,

Table 5
The Nikkei 225 opening cash prices forecasting results using random walk, SVR and ICA–SVR models.

Models	Metrics					
	RMSE	NMSE	MAD	DS	CD	CP
Random walk	137.85	0.1431	105.77	50.43%	54.26%	45.96%
SVR	60.53	0.0285	43.71	83.67%	86.17%	80.75%
ICA–SVR model	56.76	0.0226	40.86	87.53%	88.77%	86.09%
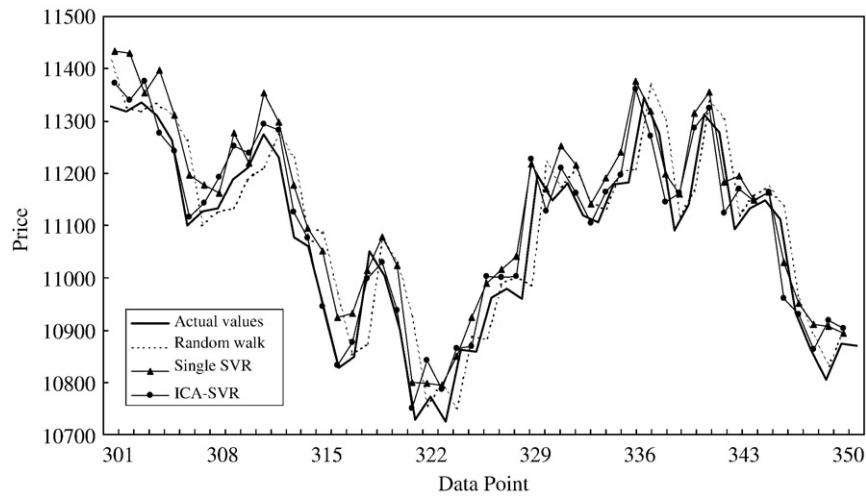
Fig. 7. The actual Nikkei 225 opening cash index and its predicted values from the random walk, SVR and ICA–SVR models, using the last 50 data points of the Nikkei 225 index in Fig. 5 as example.

NMSE and MAD of the ICA–SVR model are, respectively, 56.76, 0.0026 and 40.86. It can be observed that these values are smaller than those of random walk and SVR models. It indicates that there is a smaller deviation between the actual and predicated values using the proposed ICA–SVR model. Moreover, compared to the random walk and SVR models, the ICA–SVR model has the highest DS (directional Symmetry), CP (correct up trend) and CD (correct down trend) ratios which are 87.53%, 88.77% and 86.09%, respectively. DS, CP and CD provide a good measure of the consistency in prediction of the price direction. Thus, it can be concluded that the proposed ICA–SVR model provides a better forecasting result than the random walk and SVR models in terms of prediction error and prediction accuracy.

The actual Nikkei 225 opening cash price values and predicted values from the random walk, SVR and ICA–SVR models are illustrated in Fig. 7. Note that, to save space, the last 50 data points of the Nikkei 255 index in Fig. 5 are used as illustrative example and shown in Fig. 7. It can be observed from Fig. 7 that the predicted values obtained from the proposed ICA–SVR model are closer to the actual values than those of random walk and SVR models.

The proposed ICA–SVR method also performs well in forecasting the TAIEX closing cash prices. Table 6 summarizes the TAIEX closing cash prices forecasting results using random walk, SVR and ICA–SVR models. It can also be observed from Table 6 that the proposed ICA–SVR model has the smallest RMSE, NMSE and MAD values and the highest DS, CP and CD values in comparison with random walk and SVR models. Thus, the proposed method can produce lower prediction errors and higher prediction accuracy on the direction of change in price and outperforms random walk and SVR models in forecasting of the TAIEX closing cash prices.

Fig. 8 depicts the actual TAIEX closing cash price values and predicted values from the random walk, SVR and ICA–SVR models, using the last 50 data points of the TAIEX index in Fig. 6 as an illustrative example. From the figure, it can be observed that the proposed ICA–SVR model provides good forecasting results. The

predicted values of the proposed model are very close to the actual values than those of random walk and SVR models.

5.3. Robustness evaluation

To evaluate the robustness of the proposed ICA–SVR method, the performance of the random walk, SVR 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 Nikkei 225 opening cash index and TAIEX closing cash index by the three methods are summarized in Table 7 in terms of two criteria, RMSE and directional Symmetry (DS). Based on the findings in Table 7, it can be observed that the proposed ICA–SVR method outperforms the other benchmarking tools under all four different ratios in terms of the RMSE and DS criteria. It therefore indicates that ICA–SVR based approach indeed provides better forecasting accuracy than the other two approaches. Nevertheless, under a ratio of only 60%, the proposed ICA–SVR based approach can still provide reasonably good forecasting results (DS higher than 80%). The proposed method can effectively detect and remove the noise from financial time series data and improve the forecasting performance of SVR.

5.4. Significance test

In order to test whether the proposed ICA–SVR model is superior to the single SVR and random walk models in financial time series forecasting, the Wilcoxon signed-rank test is applied. The test is a distribution-free, non-parametric technique that does not require any underlying distributions in the data, and deals with the signs and ranks of the values and not with their magnitude (thus not influenced by outlier data points). It is one of the most commonly adopted tests in evaluating the predictive capabilities of two different models to see whether they are statistically significant difference between them [18,27,49,51,54,65].

The test procedure first calculates the differences between the paired observations, ranks them from the smallest to largest by absolute value, and then affixes the sign of each difference to the corresponding rank [49]. The sum of the ranks having a plus sign is called $J+$, and the sum of the ranks having a minus sign is called $J-$. When the sample size n is larger than 25, the distribution of J (where either $J+$ or $J-$ may be used for J) is closely approximated by a normal distribution with a mean of $u_J = \frac{n(n+1)}{4}$ and a standard

Table 6
The TAIEX closing cash prices forecasting results using random walk, SVR and ICA–SVR models.

Models	Measure					
	RMSE	NMSE	MAD	DS	CD	CP
Random walk	53.21	0.1692	39.88	46.15%	45.22%	47.93%
SVR	46.60	0.0330	34.63	55.98%	52.10%	60.60%
ICA–SVR model	41.09	0.0297	31.70	60.15%	55.94%	63.73%
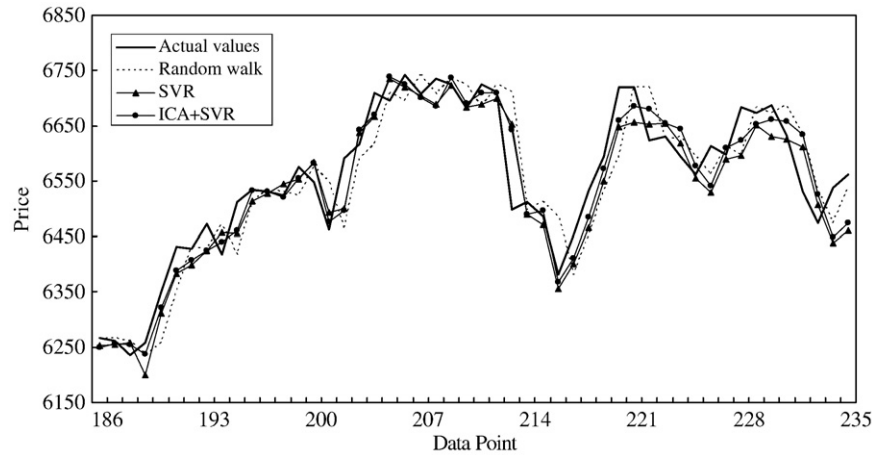
Fig. 8. The actual TAIEX closing cash index and its predicted values from the random walk, SVR and ICA–SVR models, using the last 50 data points of the TAIEX index in Fig. 6 as example.

error of $\sigma_J = \sqrt{\frac{n(n+1)(2n+1)}{24}}$. Thus the test statistic can be calculated from $Z = \frac{|J - u_J| - 0.5}{\sigma_J}$, where for J we may use, with identical results, either $J+$ or $J-$. For the details of the Wilcoxon signed-rank test, please refer to Diebold and Mariano [18] and Pollock et al. [49].

We employ the test to evaluate the predictive performance of the three built models under different ratios of the size of the training data set to complete data set. Table 8 presents the Z statistic values of the two-tailed Wilcoxon signed-rank test for RMSE values between the proposed ICA–SVR model and other two models, where the numbers in parentheses are the corresponding p-values. It can be observed from Table 8, under different ratios of training sample dataset size to the complete dataset size, that the RMSE values of the proposed ICA–SVR model is significantly different from the SVR and random models. As the proposed method can generate the smallest RMSE values in all experimental conditions of this study, we can therefore conclude that

it is significantly better than the other two models in financial time series forecasting.

6. Conclusions

This paper proposed a two-stage forecasting model by integrating ICA and SVR for financial time series. The proposed ICA–SVR method first uses ICA based on reconstruction criterion to remove the noise from forecasting variables since the financial time series data is inherently noisy. The noise in the data could lead to an over-fitting or under-fitting problem. The filtered forecasting variables containing less noise information are then used in SVR for building forecasting model. The experiments have evaluated two datasets including the Nikkei 225 opening cash index and the TAIEX closing cash index. This study compared the proposed method with traditional SVR and random walk models using prediction error and prediction accuracy as criteria. Experimental results showed that the proposed model can produce lower prediction error and higher prediction accuracy and outperformed the SVR and random walk models. According to the experiments, it can be concluded that the proposed method can effectively detect and remove the noise from financial time series data and improve the forecasting performance of SVR. Future researches may aim at combining ICA and other forecasting tools, like neural networks and grey system theory, in evaluating the ability of the proposed de-noise forecasting scheme. Integrating SVR and other signal processing techniques, like wavelet transform and nonnegative matrix factorization, in further improving the forecasting capabilities can also be investigated in future studies.

Appendix A

The *RHD* equation used in this study is as follows (Cheung and Xu [13]):

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

Table 7
Robustness evaluation of random walk, SVR and ICA–SVR models by different training and testing sample sizes.

Relative ratio(%)	Models	Nikkei 225 Testing RMSE	Nikkei 225 Testing DS(%)	TAIEX Testing RMSE	TAIEX Testing DS(%)
60	Random walk	138.54	50.21	49.85	47.58
	SVR	62.27	81.70	47.47	54.24
	ICA–SVR model	57.40	86.53	44.92	59.27
70	Random walk	137.85	50.43	53.21	46.15
	SVR	60.53	83.67	46.60	55.98
	ICA–SVR model	56.76	87.53	41.09	60.15
80	Random walk	144.43	49.12	56.78	47.09
	SVR	54.10	84.34	40.71	59.64
	ICA–SVR model	49.53	89.07	34.39	65.29
90	Random walk	132.69	53.10	61.63	37.66
	SVR	42.63	88.57	38.80	60.84
	ICA–SVR model	39.21	92.71	30.46	67.14

Table 8
Wilcoxon signed-rank test between ICA–SVR model and random walk, SVR models by different ratios of training and testing sample sizes.

Models	Relative ratio (%)	SVR		Random walk	
		Nikkei 225	TAIEX	Nikkei 225	TAIEX
ICA–SVR	60	8.62 (0.000)	6.99 (0.000)	11.43 (0.000)	7.04 (0.000)
	70	8.25 (0.000)	6.56 (0.000)	13.89 (0.000)	6.61 (0.000)
	80	6.15 (0.000)	4.07 (0.000)	14.61 (0.000)	4.65 (0.000)
	90	4.09 (0.000)	3.41 (0.000)	9.96 (0.000)	3.52 (0.000)

The numbers in parentheses are the corresponding p-values.

where $R_i = sign[T_i(t+1) - T_i(t)]$; $\hat{R}_i = sign[A_i(t+1) - A_i(t)]$;

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

T_i is actual value; A_i is predicted value; n is total number of data points.

References

[1] A. Back, A. Weigend, Discovering structure in finance using independent component analysis, Proceedings of 5th International Conference on Neural Networks in Capital Market, Kluwer Academic, 1997, pp. 15–17.

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

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

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

[5] G.E.P. Box, G.M. Jenkins, Time Series Analysis, Forecasting and Control, Holden-Day, San Francisco, 1970.

[6] R. Burbidge, M. Trotter, B. Buxton, S. Holden, Drug design by machines learning: support vector machines for pharmaceutical data analysis, Computer & Chemistry 26 (2001) 5–14.

[7] L.J. Cao, Support vector machines experts for time series forecasting, Neurocomputing 51 (2003) 321–339.

[8] L.J. Cao, F.E.H. Tay, Financial forecasting using support vector machines, Neural Computing & Applications 10 (2001) 184–192.

[9] L.J. Cao, W.K. Chong, Feature extraction in support vector machine: a comparison of PCA, XPCA and ICA, Proceedings of the 9th International Conference on Neural Information, vol. 2, 2002, pp. 1001–1005.

[10] C.C. Chang, C.J. Lin, LIBSVM: A Library for Support Vector Machines, 2001 Online available: http://www.csie.ntu.edu.tw/~cjlin/libsvm.

[11] J. Chen, Z.H. Song, P. Li, Soft sensor modeling based on DICA-SVR, Lecture Notes in Computer Science 3644 (2005) 868–877.

[12] V. Cherkassky, Y. Ma, Practical selection of SVM parameters and noise estimation for SVM regression, Neural Networks 17 (2004) 113–126.

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

[14] K.W. Cheung, J.T. Kwok, M.H. Law, K.C. Tsui, Mining customer product ratings for personalized marketing, Decision Support Systems 35 (2003) 231–243.

[15] C.C. Chuang, S.F. Su, J.T. Jeng, C.C. Hsiao, Robust support vector regression networks for function approximation with outliers, IEEE Transactions on Neural Networks 13 (2002) 1322–1330.

[16] V. David, A. Sanchez, Frontiers of research in BSS/ICA, Neurocomputing 49 (2002) 7–23.

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

[18] F.X. Diebold, R.S. Mariano, Comparing predictive accuracy, Journal of Business and Economic Statistics 13 (1995) 253–263.

[19] J.W. Hall, Adaptive selection of U.S. stocks with neural nets, in: G.J. Deboeck (Ed.), Trading on the Edge: Neural, Genetic and Fuzzy Systems for Chaotic Financial Markets, Willey, New York, 1994, pp. 45–65.

[20] S. Haykin, Neural Network: a Comprehensive Foundation, Prentice Hall, New Jersey, 1999.

[21] Z.Y. Hou, S.W. Yao, Y.Q. Gu, J.Q. Xu, Independent component analysis-support vector regression and its application in near infrared spectral analysis, Journal of Henan Normal University (Natural Science) 34 (2006) 75–78.

[22] Z. Huang, H. Chen, C.J. Hsu, W.H. Chen, S. Wu, Credit rating analysis with support vector machines and neural networks: a market comparative study, Decision Support Systems 37 (2004) 543–558.

[23] A. Hyvärinen, Fast and robust fixed-point algorithms for independent component analysis, IEEE Transactions on Neural Networks 10 (1999) 626–634.

[24] A. Hyvärinen, E. Oja, Independent component analysis: algorithms and applications, Neural Networks 13 (2000) 411–430.

[25] A. Hyvärinen, J. Karhunen, E. Oja, Independent Component Analysis, John Wiley & Sons, New York, 2001.

[26] H. Ince, T.B. Trafalis, A hybrid model for exchange rate prediction, Decision Support Systems 42 (2006) 1054–1062.

[27] T. Jaditz, L.A. Riddick, C.L. Sayers, Multivariate nonlinear forecasting using financial information to forecast the real sector, Macroeconomic Dynamics 2 (1998) 369–382.

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

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

[30] K.J. Kim, Financial time series forecasting using support vector machines, Neurocomputing 55 (2003) 307–319.

[31] K.I. Kim, K. Jung, S.H. Park, H.J. Kim, Support vector machines for texture classification, IEEE Transactions on Pattern Analysis and Machine Intelligence 24 (2002) 1542–1550.

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

[33] K. Kiviluoto, E. Oja, Independent component analysis for parallel financial time series, Proceeding of the Fifth International Conference on Neural Information, Tokyo, Japan, 1998, pp. 895–898.

[34] M. Lam, Neural network techniques for financial performance prediction: integrating fundamental and technical analysis, Decision Support Systems 37 (2004) 567–581.

[35] T.W. Lee, Independent Component Analysis: Theory and Application, Kluwer Academic Publishers, Boston, 1998.

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

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

[38] W. Leigh, R. Purvis, J.M. Ragusa, Forecasting the NYSE composite index with technical analysis, pattern recognizer, neural network, and genetic algorithm: a case study in *romantic* decision support, Decision Support Systems 32 (2002) 361–377.

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

[40] S. Li, J.T. Kwok, H. Zhu, Y. Wang, Texture classification using the support vector machines, Pattern Recognition 36 (2003) 2883–2893.

[41] C.J. Lin, C.W. Hsu, C.C. Chang, A practical guide to support vector classification, Technical Report, Department of Computer Science and Information Engineering, National Taiwan University, 2003.

[42] C.J. Lu, D.M. Tsai, Independent component analysis-based defect detection in patterned liquid crystal display surfaces, Image and Vision Computing 26 (2008) 955–970.

[43] C.J. Lu, C.M. Wu, C.J. Keng, C.C. Chiu, Integrating application of SPC/EPC/ICA and neural networks, International Journal of Production Research 46 (2008) 873–893.

[44] D. Martens, L. Bruynseels, B. Baesens, M. Willekens, J. Vanthienen, Predicting going concern opinion with data mining, Decision Support Systems 45 (2008) 765–777.

[45] M.A. Mohandes, T.O. Halawani, S. Rehmam, A.A. Hussain, Support vector machines for wind speed prediction, Renewable Energy 29 (2004) 939–947.

[46] U. Norinder, Support vector machine models in drug design: applications to transport processes and QSAR using simplex optimizations and variable selection, Neurocomputing 55 (2003) 337–346.

[47] E. Oja, K. Kiviluoto, S. Malaroiu, Independent component analysis for financial time series, Proceeding of the IEEE 2000 Adaptive Systems for Signal Processing, Communications, and Control Symposium, Lake Louise, Canada, 2000, pp. 111–116.

[48] P.F. Pai, C.S. Lin, Using support vector machines in forecasting production values of machinery industry in Taiwan, International Journal of Advanced Manufacturing Technology 27 (2005) 205–210.

[49] A.C. Pollock, A. Macaulay, M.E. Thomson, D. Onkal, Performance evaluation of judgemental directional exchange rate predictions, International Journal of Forecasting 21 (2005) 473–489.

[50] A.P. Sinha, H. Zhao, Incorporating domain knowledge into data mining classifiers: an application in indirect lending, Decision Support Systems 46 (2008) 287–299.

[51] B.L. Smith, B.M. Williams, R.K. Oswald, Comparison of parametric and nonparametric models for traffic flow forecasting, Transportation Research Part C 10 (2002) 303–321.

[52] Z.L. Sun, T.M. Choi, K.F. Au, Y. Yu, Sales forecasting using extreme learning machine with applications in fashion retailing, Decision Support Systems 46 (2008) 411–419.

[53] J.A.K. Suykens, J. De Brabanter, L. Lukas, J. Vandewalle, Weighted least squares support vector machines: robustness and sparse approximation, Neurocomputing 48 (2002) 85–105.

[54] N.R. Swanson, H. White, Forecasting economic time series using flexible versus fixed specification and linear versus nonlinear econometric models, International Journal of Forecasting 13 (1997) 439–461.

[55] F.E.H. Tay, L.J. Cao, Application of support vector machines in financial time series forecasting, Omega 29 (2001) 309–317.

[56] F.E.H. Tay, L.J. Cao, Support vector machine with adaptive parameters in financial time series forecasting, IEEE Transactions on Neural Networks 14 (2003) 1506–1518.

[57] U. Thissen, R. Van Brakel, A.P. De Weijer, W.J. Melssen, L.M.C. Buydens, Using support vector machines for time series prediction, Chemometrics and Intelligent Laboratory Systems 69 (2003) 35–49.

[58] D.M. Tsai, P.C. Lin, C.J. Lu, An independent component analysis based filter design for defect detection in low-contrast surface images, Pattern Recognition 39 (2006) 1679–1694.

[59] V.N. Vapnik, An overview of statistical learning theory, IEEE Transactions on Neural Networks 10 (1999) 988–999.

[60] V.N. Vapnik, The Nature of Statistical Learning Theory, Springer, New York, 2000.

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

[62] G. Wang, Y. Sun, Q. Ding, C. Dong, D. Fu, C. Li, Estimation of source spectra profiles and simultaneous determination of polycomponent in mixtures from ultraviolet spectra data using kernel independent component analysis and support vector regression, Analytica Chimica Acta 594 (2007) 101–106.

[63] S. Wood, Float Analysis: Powerful Technical Indicators Using Price and Volume, John Wiley & Sons, New York, 2002.

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

[65] G.P. Zhang, An investigation of neural networks for linear time-series forecasting, Computers & Operations Research 28 (2001) 1183–1202.

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

Tian-Shyug Lee is a professor in the Graduate Institute of Management at Fu-Jen Catholic University. He got his Ph.D. in Operations Research and Industrial Engineering from the University of Texas at Austin. His research and teaching interests are in the area of applied statistics and probability, application of artificial intelligence and data mining. He has published articles in various journals, including Computational Statistics and Data Analysis, Digestive and Liver Disease, Expert Systems with Applications, International Journal of Systems Science, International Journal of Fuzzy Systems, Journal of Human Resource and Adult Learning, Journal of Intelligent Manufacturing, Probability in the Engineering and Informational Sciences, and Obesity Surgery.

Chi-Jie Lu is an assistant professor in the Department of Industrial Engineering and Management at Ching Yun University, Taiwan. He got his Ph.D. in Industrial Engineering and Management from Yuan-Ze University, Taiwan. His research and teaching interests are in the area of data pattern recognition and classification, data mining, and machine vision and inspection. He has published articles in various journals, including Pattern Recognition, Image and Vision Computing, International Journal of Production Research, Computational Statistics and Data Analysis, and Expert Systems with Applications.

Chih-Chou Chiu is a professor in the Institute of Commerce Automation and management at National Taipei University of Technology. His Ph.D. is from Texas A&M University in Industrial Engineering. His research and teaching interests are in the area of application of artificial intelligence, Bayesian statistical approach, data mining and continuous process improvement techniques for manufacturing. He has published articles in various journals, including IIE Transactions, International Journal of Production Research, Journal of Intelligent Manufacturing, International Journal of Systems Science, Quality and Reliability Engineering International, and Expert Systems with Applications.