



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

Chi-Jie Lu<sup>a</sup>, Tian-Shyug Lee<sup>b,\*</sup>, Chih-Chou Chiu<sup>c</sup>

<sup>a</sup> Department of Industrial Engineering and Management, Ching Yun University, Taiwan, ROC

<sup>b</sup> Graduate Institute of Management, Fu Jen Catholic University, Taiwan, ROC

<sup>c</sup> Institute of Commerce Automation and Management, National Taipei University of Technology, Taiwan, ROC

## ARTICLE INFO

### Article history:

Received 28 April 2008

Received in revised form 31 January 2009

Accepted 8 February 2009

Available online 13 February 2009

### Keywords:

Independent component analysis

Support vector regression

Financial time series forecasting

Stock index

## 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  $\epsilon$ -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](mailto:036665@mail.fju.edu.tw), [mleettw@yahoo.com](mailto:mleettw@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  $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_m]^T$  be a multivariate data matrix of size  $m \times n$ ,  $m \leq n$ , consisting of observed mixture signals  $\mathbf{x}_i$  of size  $1 \times n$ ,  $i = 1, 2, \dots, m$ . In the basic ICA model, the matrix  $\mathbf{X}$  can be modeled as [24]

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

متن کامل مقاله

دریافت فوری ←

**ISI**Articles

مرجع مقالات تخصصی ایران

- ✓ امکان دانلود نسخه تمام متن مقالات انگلیسی
- ✓ امکان دانلود نسخه ترجمه شده مقالات
- ✓ پذیرش سفارش ترجمه تخصصی
- ✓ امکان جستجو در آرشیو جامعی از صدها موضوع و هزاران مقاله
- ✓ امکان دانلود رایگان ۲ صفحه اول هر مقاله
- ✓ امکان پرداخت اینترنتی با کلیه کارت های عضو شتاب
- ✓ دانلود فوری مقاله پس از پرداخت آنلاین
- ✓ پشتیبانی کامل خرید با بهره مندی از سیستم هوشمند رهگیری سفارشات