CrossMark

Applying the Hybrid Model of EMD, PSR, and ELM to Exchange Rates Forecasting

Heng-Li Yang[1] · **Han-Chou Lin[1]**

Abstract Financial time series forecasting has been a challenge for time series analysts and researchers because it is noisy, nonstationary and chaotic. To overcome this limitation, this study uses empirical mode decomposition (EMD) and phase space reconstruction (PSR) to assist in the task of financial time series forecasting. In addition, we propose an approach that combines these two data preprocessing methods with extreme learning machine (ELM). The approach contains four steps as follows. (1) EMD is used to decompose the dynamics of the exchange rate time series into several components of intrinsic mode function (IMF) and one residual component. (2) The IMF and residual time series phase space is reconstructed to reveal its unseen dynamics according to the optimum time delay τ and embedding dimension m. (3) The reconstructed time series datasets are divided into two datasets: training and testing, in which the training datasets are used to build ELM models. (4) A regression forecast model is set up for each IMF as well as the residual component by using ELM. The final prediction results are obtained by compositing the prediction values. To verify the effectiveness of the proposed approach, four exchange rates are chosen as the forecasting targets. Compared with some existing state-of-the-art models, the proposed approach yields superior results. Academically, we demonstrated the validity and superiority of the proposed approach that integrates EMD, PSR, and ELM. Corporations or individuals can apply the results of this study to acquire accurate exchange rate information and reduce exchange rate expenses.

✉ Heng-Li Yang
yanh@nccu.edu.tw

Han-Chou Lin
actuallife@hotmail.com

[1] Department of Management Information Systems, National Chengchi University, Taipei, Taiwan

1 Introduction

An exchange rate reflects relative values between different currencies, and is one of the most important financial and macroeconomic indicators in an economy. The fluctuation of exchange rates affects international trade, capital flows, and asset portfolio management. Many financial time series forecasting models (Adhikari and Agrawal 2013; Wu and Chang 2012; Zhiqiang et al. 2012) have been developed that play a critical role in the world economy because of their ability to forecast economic benefits and influence economic development. These models have attracted increased attention from academic researchers and business people for its theoretical possibilities and practical applications (Hadavandi et al. 2010; Lu et al. 2009). The ostensible purpose of the breakdown of financial market boundaries was to enhance the efficiency of capital funding (for example, the Bretton Woods system of monetary management was officially ended in the 1973). As a result, currencies that are traded internationally have become crucial economic indices for international trade, financial markets, alignment of economic policy by governments, and corporate financial decision-making.

However, financial time series forecasting is a challenging task because of its inherent nonlinearity and nonstationary characteristics. In the last few decades, these characteristics have attracted increased attention from many academic researchers. The forecasting approaches used in literature can be classified into two types of models: statistical and artificial intelligence (Wang et al. 2012; Zhu and Wei 2013). Linear statistical models such as exponential smoothing (Lemke and Gabrys 2010) and autoregressive integrated moving average (ARIMA) (GEP and GM 1970) have identified immense applications for forecasting financial data. A subclass ARIMA model, namely, Naïve random walk (RW) (Sun 2005; Tyree and Long 1995), has become the benchmark statistical technique in this domain. In a simple RW model, each forecast is assumed to be the sum of the most recent observation and a random error term. After the pioneering work of Meese and Rogoff (1983), the RW model has been extensively used by many researchers for foreign exchange rate forecasting. Currently, the simple RW is the most dominant linear model in literature of the financial time series and, especially, exchange rates (Zhang 2003).

Despite the simplicity and notable forecasting accuracies of RW models, their main drawback is their inherent linear form. Thus, such statistical models cannot effectively capture nonlinear patterns hidden in financial time series because these models are developed based on the assumption that the time series being forecasted are linear and stationary (Huang et al. 2010). To overcome this limitation of statistical models, several nonlinear models have been proposed. Among them, the artificial neural network (ANN) has attracted considerable interest from researchers because of their excellent nonlinear modeling capability (Zhang and Wu 2009; Chen et al. 2012a; Jaeger and Haas 2004; Deng et al. 2015; Vasilakis et al. 2013). Many studies have concluded that the ANN model outperforms conventional statistical models. However, ANN suffers from local minimum traps and has difficulty determining hidden layer size and learning

rate (Kazem et al. 2013). A new learning algorithm for the single hidden layer feed-forward neural network (SLFN) known as the extreme learning machine (ELM) has been proposed that overcomes the aforementioned disadvantages (Huang et al. 2006a; Chen and Ou 2011). In the learning process of ELM, the input weights and hidden biases are randomly selected, and the output weights are analytically determined by using the Moore-Penrose generalized inverse. ELM can learn much faster and has a higher generalization performance than do the traditional gradient-based learning algorithms. In addition, ELM solves the problems of stopping criteria, learning rates, learning epochs, and local minima (Huang et al. 2006; Chen and Ou 2011; Xia et al. 2012; Lu and Shao 2012). In recent years, ELM has attracted considerable attention and become an important method in nonlinear modeling (Chen and Ou 2011; Xia et al. 2012; Lu and Shao 2012).

When building intelligent prediction models directly using original values, obtaining satisfactory forecast results is difficult because of the high-frequency, nonstationary, and chaotic properties of financial data. Hence, to further improve prediction performance, recent research efforts on modeling time series with complex nonlinearity, dynamic variation, and high irregularity have initially used information extraction techniques to extract features hidden in the data. They then use these extracted characteristics to construct a forecasting model (Lu et al. 2009; Chen et al. 2012b; Liu and Wang 2011; Lu 2010). In other words, by means of suitable feature extractions or signal processing methods, useful or interesting information that may not be observed directly from the original data can be revealed in the extracted features. Therefore, an effective forecasting model possessing more precise prediction capabilities must be developed.

Empirical mode decomposition (EMD), based on Hilbert-Huang transform (HHT), is suitable for decomposing nonlinear and nonstationary time series, which adaptively represent the local characteristics of the given signal (Huang et al. 1998, 2003).Though the use of EMD, any complicated signal can be decomposed into a finite and often small number of intrinsic mode functions (IMFs). IMFs possess simple frequency components and strong correlations, and thus are easy and accurate to forecast (Jaeger and Haas 2004). EMD has been widely used in many fields, including in the analysis of the atmosphere time series (Xuan and Yang 2008), river water turbidity forecasting (Wang and Qi 2009), crude oil price prediction (Yang et al. 2010), short-term wind power prediction, and others (Jaeger and Haas 2004; Lu and Shao 2012; Chen et al. 2012a; Bao et al. 2012; Ye and Liu 2011).

Another critical reason that financial time series are notoriously difficult to predict is their chaotic nature. Chaos is often identified in the fields of physics and other natural sciences. Empirical evidence of chaotic behavior in financial time series has also been identified (Barkoulas and Travos 1998; Gimore 2001; McKenzie 2001). Chaos theory points out that an adequate method can help reveal underlying information in complicated matters believed to be unpredictable (Takens 1981). For chaotic time series, the techniques of prediction based on phase space reconstruction (PSR) can be employed to extract information and characteristics hidden in dynamic systems of time series. PSR can transform a one-dimensional signal into a structure that embeds sufficiently high dimensions. In this new high dimensional space, a structure is formed that is

topologically equivalent to the original phase space. This has led some researchers to apply chaos theory to time series forecasting.

In this study, we propose a hybrid exchange rate forecasting model by integrating EMD, PSR, and ELM (EMD + PSR + ELM). First, the original exchange rate time series are first decomposed into a finite number of independent IMFs employing different frequencies. Second, based on PSR, different ELM models are used to model and forecast the four sub-series, respectively, according to reconstructed time series. Finally, these forecasting results are combined with the ultimate forecasting result output. Moreover, experimental results from four sets of real exchange rate data demonstrate that the proposed hybrid forecasting method outperforms methods of Naïve RW, single ELM, and other hybrid models in terms of mean absolute error (MAE), root mean-square error (RMSE), and mean absolute error (MAPE).

2 Literature Review of Major Methods

2.1 EMD

The EMD method based on HHT is based on the simple assumption that any signal consists of different but simple intrinsic mode oscillations. The essence of the method is to identify the intrinsic oscillatory modes (IMFs) (Huang et al. 1998) based on their characteristic time scales in the signal and then decompose the signal accordingly. A characteristic time scale is defined by the time lapse between the successive extremes.

To extract the IMF from a given data set, the sifting process is implemented as follows. First, we identify all local extrema, and then connect all local maxima by a cubic spline line that thus acts as the upper envelope. Then, we repeat the procedure for the local minima to produce the lower envelope. The upper and lower envelopes should cover all the data between them. Their mean is designated $m_1(t)$, and the difference between the data and $m_1(t)$ is $h(t)$, given by the following:

$$x(t) - m_1(t) = h_1(t). \tag{1}$$

Ideally, $h_1(t)$ should be an IMF. Because the construction of $h_1(t)$ described previously should have forced the result to satisfy all definitions of an IMF, we demand the following conditions: (i) $h_1(t)$ should be free of riding waves, that is, the first component should not display under- or over-shots that ride on the data and produce local extremes without zero crossing; (ii) symmetry of the upper and lower envelops with respect to zero should be displayed; (iii) the number of zero crossing and extremes should be the same in both functions.

The sifting process must be repeated as many times as required to reduce the extracted signal to an IMF. In the subsequent sifting process steps, $h_1(t)$ is treated as the data:

$$h_1(t) - m_{11}(t) = h_{11}(t), \tag{2}$$

where $m_{11}(t)$ is the mean of the upper and lower envelops of $h_s(t)$. This process can be repeated as many as k times and $h_{1k}(t)$ is then defined as:

$$h_{1(k-1)}(t) - m_{1k}(t) = h_{1k}(t). \tag{3}$$

After each processing step, we must confirm that the number of zero crossings equals the number of extrema. The resulting time series is the first IMF and then is designated as $c_1(t) = h_{1k}(t)$. The first IMF component from the data contains the highest oscillation frequencies found in the original data $x(t)$.

This first IMF is subtracted from the original data, and this difference is called a residue $r_1(t)$ by means of the following:

$$x(t) - c_1(t) = r_1(t). \tag{4}$$

The residue $r_1(t)$ is considered as if it was the original data and we reapply the sifting process to it. The process of locating additional intrinsic modes $c_1(t)$ continues until the last mode is found. The final residue will be a constant or a monotonic function. In this last case, it will be the general trend of the data.

$$x(t) = \sum_{j=1}^{n} c_j(t) + r_n(t). \tag{5}$$

Thus, the data is decomposed into n-empirical IMF modes plus a residue, $r_n(t)$, which can be either the mean trend or a constant.

2.2 Phase Space Reconstruction

The analysis of time series generated by non-linear dynamic systems can be accomplished in accordance with Takens' embedding theory (Takens 1981). Given a univariate time series $\{x_i\}_{i=1}^{N}$ generated from a d-dimension chaotic attractor and where N is the length of the time series, a phase space R^d of the attractor can be reconstructed by using a delay coordinate defined as:

$$X_i = (x_i, x_{i-\pi}, \ldots, x_{i-(m-1)\pi}), \tag{6}$$

where m is both the embedding dimension of reconstructed phase space and the time delay constant. Choosing the correct embedding dimension is crucial for predicting x_{t+1}. Takens (Takens 1981) considered that a sufficient condition for the embedding dimension is $m \geq 2d + 1$. However, an embedding dimension that is too large requires additional observations and complex computation. Moreover, if we choose an embedding dimension that is too large, noise and other unwanted inputs will be embedded with the real source input information. This may then corrupt the underlying system dynamic information. Therefore, in accordance with Sauer et al. (1991), if the dimension of the original attractor is d, then an embedding dimension of $m = 2d + 1$ is adequate for reconstructing the attractor.

An efficient method of locating the minimal sufficient embedding dimension is the false nearest neighbors (FNN) procedure proposed by Kennel et al. (1992). Two near points in reconstructed phase space are called false neighbors if they are considerably far apart in the original phase space. Such a phenomenon occurs if we select an embedding dimension that is lower than the minimal sufficient value and if the reconstructed attractor does not therefore preserve the topological properties of the real phase space. In this case, points are projected into the false neighborhood of other points. The idea behind the FNN procedure is as follows. Suppose X_i has a nearest neighbor X_j in an m-dimensional space. Calculate the Euclidean distance $||X_i - X_j||$ and compute the following:

$$R_i = \frac{\left\| X_{i+1} - X_{j+1} \right\|}{\left\| X_i - X_j \right\|}. \tag{7}$$

If R_i exceeds a given threshold R_{tol} (say, 10 or 15), the point X_j is considered a false nearest neighbor in dimension m. We can say that the embedding dimension m is sufficiently high if the fraction of points that have false nearest neighbors is zero or considerably small.

Estimation of time delay τ is another major concern. If τ is too small, redundancy will occur. However, if τ is too large, it will probably lead to a complex phenomenon called irrelevance. In this study, we use the first minimum of mutual information (MI) function (Huang et al. 2006a) to determine τ as follows:

$$MI(\tau) = \sum_{n=1}^{N-\pi} P(x_n, x_{n+\pi}) \log_2 \left(\frac{P(x_n, x_{n+\pi})}{P(x_n) P(x_{n+\pi})} \right), \tag{8}$$

where $P(x_n)$ is the probability density of x_n and $P(x_n, x_{n+\tau})$ is the joint probability density of x_n and $x_{n+\tau}$.

2.3 ELM

ELM is an improved learning algorithm for the SLFN architecture. ELM is different from the traditional neural network methodology in that all the parameters of the feed-forward networks (input weights and hidden layer biases) are not required to be tuned. The ability of SLFNs to choose input weights randomly, as well as hidden layer biases and a nonzero activation function to approximate any continuous functions on any input set, has been demonstrated in Rao and Mitra (1971). The SLFN with randomly chosen input weights and hidden layer biases can be considered a linear system. For this linear system, the output weights that link the hidden layer to the output layer can be analytically determined through a simple generalized inverse operation of the hidden layer output matrices. This simple approach enables ELM to be extremely efficient and many times faster than the traditional feed-forward learning algorithms.

The structure of ELM consists of an SLFN in which the input weight matrix W is randomly chosen and the output weight matrix β is analytically determined.

Suppose we are given a data set with N arbitrary distinct samples (x_i, t_i), where $x_i = [x_{i1}, x_{i2}, ..., x_{i1}]^T 2 \in R^n$ and $t_i = [t_{i1}, t_{i2}, ..., t_{im}]^T \in R^m$. The mathematical model of a standard SLFN with $\tilde{N}$ hidden nodes and activation function $g(x)$ for the given data can be formulated as follows (Huang et al. 2006):

$$\sum_{i=1}^{\tilde{N}} \beta_i g_i(x_j) = \sum_{i=1}^{\tilde{N}} \beta_i g_i(w_i x_j + b_i) = y_j, \quad j = 1, \ldots, N, \qquad (9)$$

where $w_i = [w_{i1}, w_{i2}, ..., w_{in}]^T$ denotes the weight vector that connects the input nodes to the ith hidden node and $b_i = [b_{i1}, b_{i2}, ..., b_{im}]^T$ is the weight vector that connects the output nodes with the ith hidden node. In addition, b_i is the threshold of the ith hidden node. The inner product of w_i and x_j is denoted by the operation $w_i \cdot x_j$ in (9). Let us consider that standard SLFNs with $\tilde{N}$ hidden nodes employing activation function $g(x)$ can approximate these N samples with zero error. In such a situation, we obtain the following equation:

$$\sum_{j=1}^{N} \| y_j - t_j \| = 0, \qquad (10)$$

where y denotes the actual output value of the SLFN. This indicates the existence of β_i, w_i, and b_i such that:

$$\sum_{i=1}^{\tilde{N}} \beta_i g_i(w_i x_j + b_i) = t_j, \quad j = 1, \ldots, N. \qquad (11)$$

A succinct expression of the previous N equations can be written as:

$$H\beta = T, \qquad (12)$$

where H is the hidden layer output matrix.

$$H = \begin{bmatrix} h(x_1) \\ \vdots \\ h(x_N) \end{bmatrix} = \begin{bmatrix} h_1(x_1) & \cdots & h_{\tilde{N}}(x_1) \\ \cdots & \cdots & \cdots \\ h_1(x_N) & \cdots & h_{\tilde{N}}(x_N) \end{bmatrix}, \qquad (13)$$

$$\beta = \begin{bmatrix} \beta_1^T \\ \vdots \\ \beta_{\tilde{N}}^T \end{bmatrix}, \qquad (14)$$

$$T = \begin{bmatrix} T_1^T \\ \vdots \\ T_N^T \end{bmatrix}. \qquad (15)$$

As previously discussed, the input weights and hidden biases are randomly generated and do not require any tuning as in the case with traditional SLFN methodology. The evaluation of the output weights that link the hidden layer to the output layer is equivalent to determining the least-square solution to the given linear system. The minimum norm least-square (LS) solution to the linear system defined in (12) is:

$$\hat{\beta} = H^+T. \tag{16}$$

The H^+ in the previous equation is the Moore–Penrose (MP) generalized inverse of matrix H (Babovic et al. 2000). The minimum norm LS solution is unique and has the smallest norm among all the LS solutions. The MP inverse-method-based ELM is shown to achieve a quality generalization performance with a radically increased learning speed. A general algorithm for ELM can be stated as follows. For a given training set, including activation function $g(x)$ and hidden neuron number L:

Step 1: Assign random input weight w_i and bias b_i, $i = 1, ..., L$.
Step 2: Calculate the hidden layer output matrix H.
Step 3: Calculate the output weight $\beta : \beta = H^+T$.

3 Proposed Model

The proposed hybrid approach for exchange rate forecasting (EMD-PSR-ELM) combines EMD, PSR, and ELM, and consists of four main stages. These four stages are described as follows.

Stage 1 EMD Decomposition

The original time series $x(t)$, $t = 1, 2, \ldots, N$ is decomposed into n IMF components, $c_j(t)$, $j = 1, 2, \ldots, n$, and one residual component $r_n(t)$ by using EMD.

Stage 2 Phase Space Reconstruction

First, the MI function in (8) is calculated for each $c_j(t)$ and $r_n(t)$ time series. Second, the first delay time in which the MI function minimum value occurs is considered the optimum time delay τ. Third, the FNN method is employed to find the minimum sufficient embedding dimension m. Fourth, according to the optimum time delay τ and embedding dimension m, the time series phase space is reconstructed to reveal its unseen dynamics.

Therefore, the input and output samples can be represented by the matrix X and Y, respectively, in the following forms (where x can denote c_j and r_n):

$$X = \begin{bmatrix} x(1) & x(1+\tau) & \cdots & x(1+(m-1)\tau) \\ x(2) & x(2+\tau) & \cdots & x(2+(m-1)\tau) \\ \vdots & \vdots & \vdots & \vdots \\ x(M) & x(M+\tau) & \cdots & x(M+(m-1)\tau) \end{bmatrix}, \quad Y = \begin{bmatrix} x(1+(m-1)\tau+lag) \\ x(2+(m-1)\tau+lag) \\ \vdots \\ x(M+(m-1)\tau+lag) \end{bmatrix}. \tag{17}$$
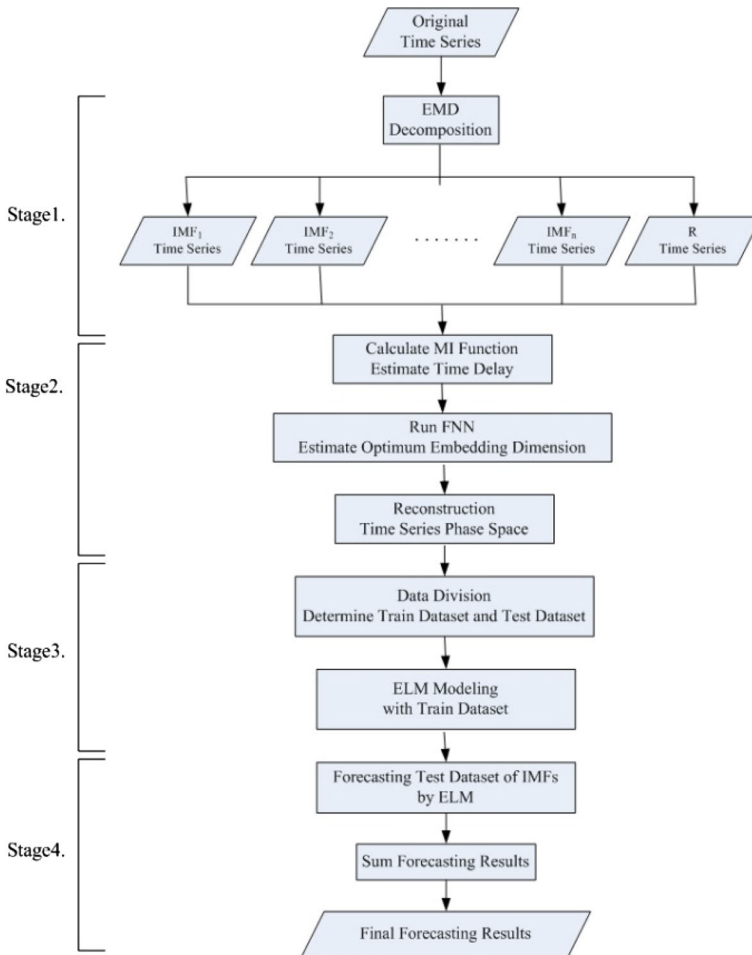
Fig. 1 Flow chart of the proposed EMD-PSR-ELM Model

Some forecasting techniques for chaotic time series nearly fix their selected time *lag* at 1 (Liong and Sivapragasam 2002; Makridakis 1993). Therefore, for this study, we also fix time *lag* = 1.

Stage 3 ELM Modeling

The reconstructed time series datasets are divided into training and testing datasets. The training datasets are used to build ELM models.

Stage 4 Result Composition

A regression forecast model is set up for each IMF whereas the residue is set up by using ELM. The final prediction results are obtained by compositing the prediction values. F is the ELM predictor function. The final forecasting result is $\sum_{j=1}^{n-1} F_j(c_j(t)) + F_n(r_n(t))$.

The flow chart of the proposed EMD-PSR-ELM model is shown in Fig. 1.

4 Experimental Results and Analysis

4.1 Data Sets

Daily exchange rate values for USD/TWD, EUD/TWD, GBP/TWD, and AUD/TWD were extracted from the data stream provided by OANDA (http://www.oanda.com) and used in this study. The entire data set covers the period from January 1, 2007 to December 31, 2013, yielding a total of 2557 observations. The data set was divided into two sets, training and testing data. The daily data from January 1, 2007 to April 19, 2013, generating a total of 2301 observations, were used as the training data set. Others of the daily data from April 20, 2013 to December 31, 2013, producing a total of 256 observations, were used as the testing data set. In the next section, we explain the manner in which we implement our EMD-PSR-ELM model.

4.2 Benchmark Prediction Models

As mentioned in Sect. 1, this study adopts the Naïve RW, ELM, EMD-ELM, and PSR-ELM as the benchmarks for the experiment.

(1) *Naïve RW* the Naïve RW simply takes the forecast for the next value from the current value. Thus, no fitting process is required.
(2) *ELM* the original time series $x(t)$ are directly used to build ELM models and to forecast final results. The function can be expressed as $\hat{x}(t + 1) = F(x(t))$, F refers to the ELM predictor function.
(3) *EMD-ELM* first, the original time series are decomposed by EMD into several IMF time series and one residual time series. These decomposed datasets are then utilized to build the ELM models previously mentioned into the EMD-ELM models.
(4) *PSR-ELM* we use the PSR method to reconstruct the original time series space, from which we can obtain optimum embedded dimension m and delay time τ. The reconstruction datasets are adopted to build ELM models as well. The function can be expressed as the following:

$$\hat{x}(t + 1) = F(x(t), x(t - \tau), ..., x(t - (m - 1) \cdot \tau)). \tag{18}$$

4.3 Evaluation Criteria

To evaluate the forecasting performance of the proposed model, we adopt the MAE, RMSE, and MAPE. These measures are defined as follows:

$$\text{MAE} = N^{-1} \sum_{t=1}^{N} \left| Y_{(t)} - \hat{Y}_{(t)} \right|, \tag{19}$$

$$\text{RMSE} = \left(N^{-1} \sum_{t=1}^{N} (Y_{(t)} - \hat{Y}_{(t)})^2 \right)^{1/2}, \tag{20}$$

$$\text{MAPE} = N^{-1} \sum_{t=1}^{N} \left| (Y_{(t)} - Y_{(t)})/\hat{Y}_{(t)} \right|, \tag{21}$$

where $Y_{(t)}$ and $\hat{Y}_{(t)}$ are the actual and prediction values, respectively, at time t, and N is the sample size. Note that MAE, RMSE, and MAPE are the measures of the deviation between actual and prediction values. Therefore, improved forecasting performance occurs when the values of these measures are small. However, if the results are not consistent among these criteria, we choose MAPE as suggested by Makridakis (1993) as the benchmark because MAPE is relatively more stable than are other criteria.

4.4 Implementation of EMD

Based on the previous steps described in Sect. 3, we conducted prediction experiments. First, using the EMD technique, the four exchange rate series (USD/TWD, EUD/TWD, GBP/TWD, AUD/TWD) were decomposed into 10 IMFs (IMF1–IMF10) and one residual (Residual), as shown in Fig. 2. All the extracted IMF components are graphically illustrated in the order in which they were extracted. The order of frequency (or period) from the highest frequency to the lowest is indicated. The last component is the residual of sifting. This generally represents the trend of the time series. In this study, EMD components were obtained by using the HHT MATLAB program (http://rcada.ncu.edu.tw/research1_clip_program.htm).

4.5 Implementation of PSR

In the PSR stage, MI was used to select the optimal delay time τ, which was selected based on the first minimum value of the MI function. After the optimal τ was selected, FNN (was then used to extract the minimum embedding dimension. Table 1 shows the optimal m and τ for each IMF and the residual. These optimal embedding dimensions and delay times are used to construct the input matrix (X). The data were fed to ELM forecast models and set up for each IMF and the residual. The final prediction results of the EMD-PSR-ELM model were obtained by compositing (i.e., combining separate prediction values into one value). We used Hao Cheng's Fractal MATLAB toolbox to implement the MI and FNN functions.

4.6 Forecasting Results and Analysis

To compare the performance of different models, we first applied the benchmarks, Naïve RW, ELM, PSR + ELM, and EMD + ELM, to forecast the four exchange rates, respectively. The performance comparison of five models (Naïve RW, ARIMA, back propagation neural network (BPNN), ELM, PSR + ELM, EMD + ELM, and the EMD + PSR + ELM) according to three evaluation criteria (MAPE, MAE and RMSE) is reported in Table 2. Relative errors defined as "the ratio of error to the actual value" of the five models are shown in Fig. 3.
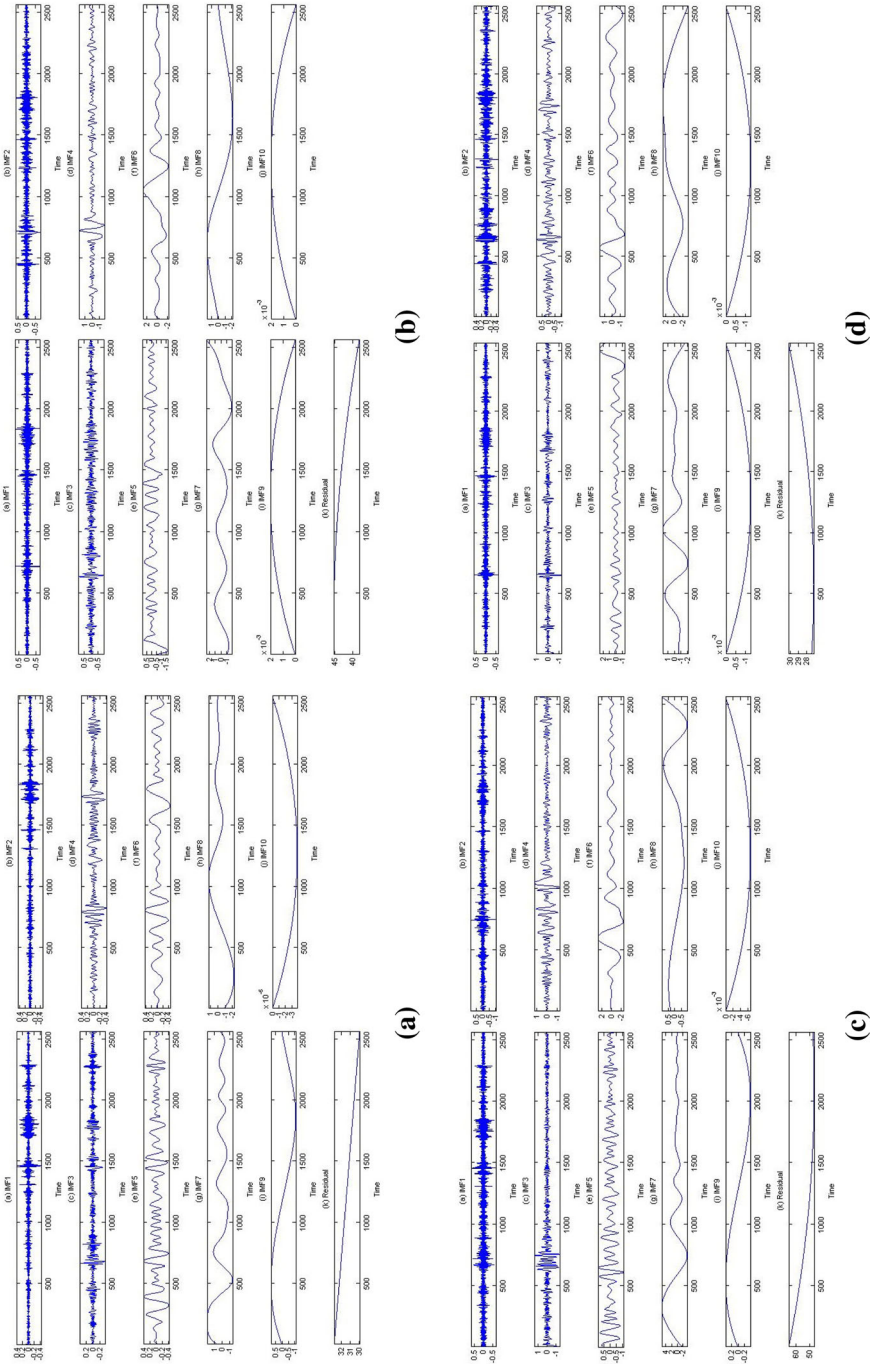
Fig. 2 IMFs and residual. a USD/TWD. b EUD/TWD. c GBP/TWD. d AUD/TWD

Table 1 Optimal m and τ for each IMF and residual

	USD/TWD		EUR/TWD		GBP/TWD		AUD/TWD	
	m	τ	m	τ	m	τ	m	τ
IMF1	7	5	7	5	6	5	5	3
IMF2	8	2	8	2	6	2	5	3
IMF3	5	4	5	4	5	7	5	5
IMF4	4	9	4	9	4	6	4	11
IMF5	3	7	3	7	3	9	3	7
IMF6	2	9	2	9	3	10	2	9
IMF7	2	9	2	9	2	9	2	7
IMF8	2	9	2	9	2	9	2	9
IMF9	1	9	1	9	2	9	1	9
IMF10	1	9	1	9	1	9	1	9
Residual	2	9	2	9	1	9	2	9

The empirical analysis confirms that the performance of EMD + PSR + ELM is the best among the five models with respect to the four exchange rates. The empirical results demonstrated the usefulness of the two-stage data preprocessing (stage 1 EMD, stage 2 PSR) of the ELM model we proposed. We can observe some phenomenon in Fig. 3 to identify possible superiority. In the high-frequency points, relative errors of the hybrid model are much smaller than in other models. This observation demonstrates that the EMD method can reduce noise contained in time series and can thus enhance accuracy.

The average error of pure ELM was the worst for MAPE, MAE, and RMSE in the four exchange rates. It was even worse than the Naïve RW in nearly all measures. This indicates that single ELM is unsuitable for exchange rate time series forecasting. However, if we combine the EMD data processing method into a single ELM model, that is, EMD + ELM model, its performance would improve.

Table 2 reveals that the accurate rates in the PSR + ELM model for the four exchange rates are not better than those in the single ELM model, even is the worst in AUD/TWD exchange rate dataset. The optimal embedding dimension m and delay time τ for the four exchange rates by using the PSR method are presented in the Tables ($m = 1, \tau = 7$), ($m = 1, \tau = 7$), ($m = 1, \tau = 7$) and ($m = 6, \tau = 7$). When $m = 1$, the input matrix (X) constructed is the same as that does not use the PSR method. Therefore, the performance for the three exchange rates, USD/TWD, EUD/TWD, and GBP/TWD, is the same as that presented in Table 2. We have no sufficient evidence to demonstrate that PSR does not enhance accuracy for exchange rate forecasting.

Furthermore, to demonstrate the effectiveness of the EMD + PSR + ELM model, we also compared ARIMA and BPNN, the most popular neural network modules, to our model. Experimental results also revealed that our model outperforms these other two with respect to MAPE, MAE, and RMSE criteria for the four data sets. This information is also shown in Table 2. It proved the strong robustness of our proposed hybrid model. Optimal parameters for ARIMA, BPNN, and PSR + ELM for the four data sets are shown in Table 3.

Table 2 Model performance for different exchange rate datasets

	Naïve random walk	Model					
		ARIMA	BPNN	ELM	PSR + ELM	EMD + ELM	EMD + PSR + ELM
	Under USD/TWD exchange rate dataset						
MAPE	0.0010292^{2}	0.0013777^{7}	0.0011106^{6}	0.0010673^{3}	0.0010673^{3}	0.0010825^{5}	0.0010251^{1}
MAE	0.0306412^{2}	0.0410707^{7}	0.0333726^{6}	0.0318183^{3}	0.0318183^{3}	0.0322375^{5}	0.0305321^{1}
RMSE	0.0613524^{4}	0.0641517^{7}	0.0611073^{3}	0.0615085^{5}	0.0615085^{5}	0.0477502^{2}	0.0425551^{1}
	Under EUD/TWD exchange rate dataset						
MAPE	0.0017743^{3}	0.0019207^{7}	0.0018426^{6}	0.0018075^{5}	0.0018024^{4}	0.0017272^{2}	0.0014101^{1}
MAE	0.0703563^{3}	0.0762317^{7}	0.0731326^{6}	0.0715094^{4}	0.0715094^{4}	0.0685212^{2}	0.0559240^{1}
RMSE	0.1164426^{6}	0.1168157^{7}	0.1163523^{3}	0.1164014^{4}	0.1164014^{4}	0.0946112^{2}	0.0747146^{1}
	Under GBP/TWD exchange rate dataset						
MAPE	0.0017743^{3}	0.0022296^{6}	0.0023567^{7}	0.0018024^{4}	0.0018024^{4}	0.0017272^{2}	0.0014101^{1}
MAE	0.0703563^{3}	0.1041546^{6}	0.1097927^{7}	0.0715094^{4}	0.0715094^{4}	0.0685212^{2}	0.0559240^{1}
RMSE	0.1164425^{5}	0.1572887^{6}	0.1581857^{7}	0.1164013^{3}	0.1164013^{3}	0.0946112^{2}	0.0747146^{1}
	Under AUD/TWD exchange rate dataset						
MAPE	0.0028393^{3}	0.0028644^{4}	0.0030357^{7}	0.0028715^{5}	0.0029826^{6}	0.0027602^{2}	0.0026431^{1}
MAE	0.0792803^{3}	0.0798524^{4}	0.0845967^{7}	0.0800795^{5}	0.0831366^{6}	0.0770882^{2}	0.0736801^{1}
RMSE	0.1177424^{4}	0.1165663^{3}	0.1191156^{6}	0.1178805^{5}	0.1193627^{7}	0.1034222^{2}	0.0959811^{1}

Superscript numbers mean the rank for the seven models under different exchange rate datasets
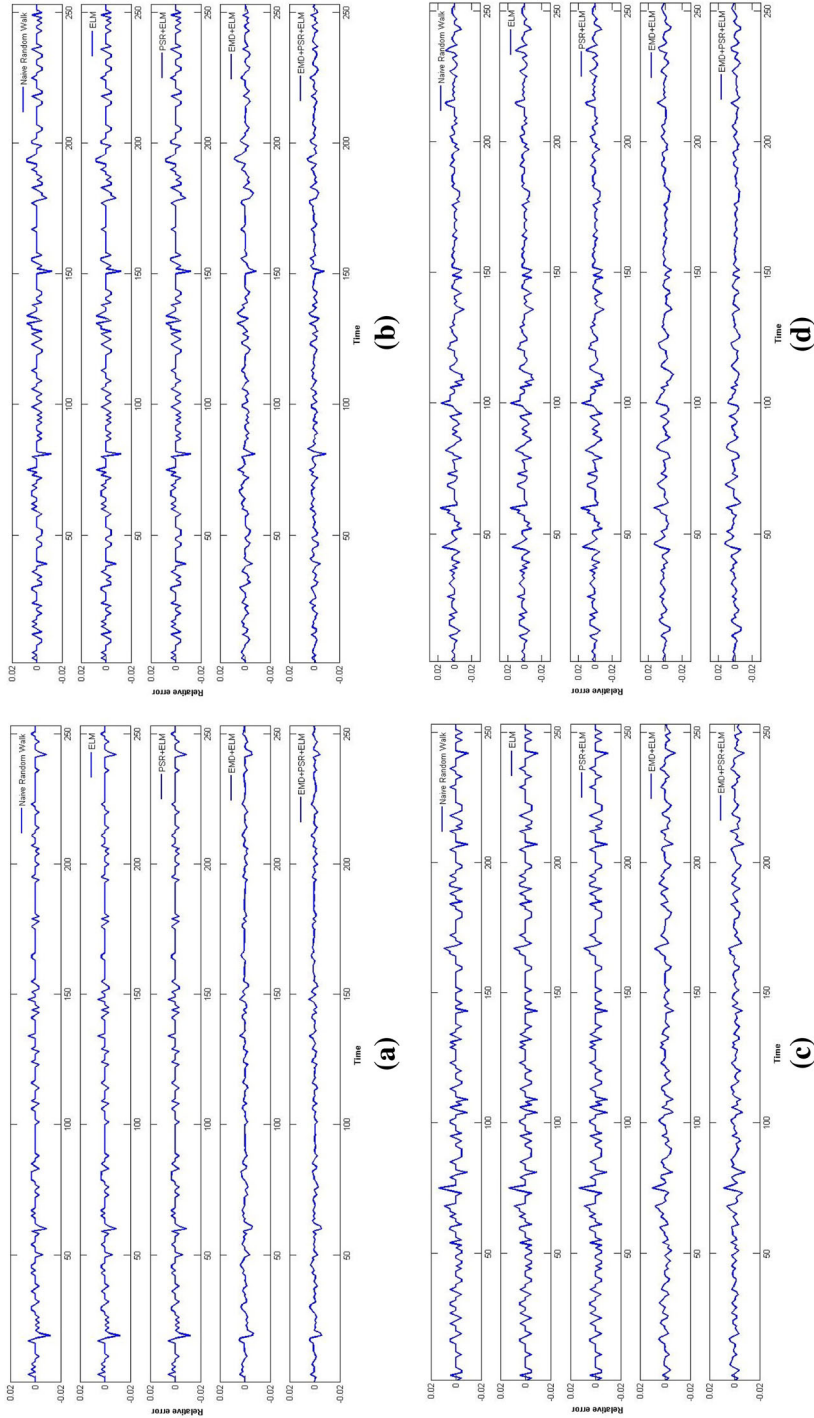
Fig. 3 Corresponding relative errors by difference models. **a** USD/TWD. **b** EUD/TWD. **c** GBP/TWD. **d** AUD/TWD

Table 3 Optimal parameter for ARIMA, BPNN, PSR + ELM

Exchange rate	Optimal parameter		
	ARIMA (p, q, r)	BPNN (hidden nodes, epochs)	PSR + ELM (m, τ)
USD/TWD	(4, 1, 4)	(2, 100)	(1, 7)
EUD/TWD	(4, 1, 4)	(1, 200)	(1, 7)
GBP/TWD	(3, 1, 3)	(1, 200)	(1, 7)
AUD/TWD	(3, 1, 1)	(4, 200)	(6, 7)

For optimal parameters of the BPNN model, 1, 2, 4 hidden nodes and 50, 100, 200, 400 epochs are employed for each stopping criterion. The learning rate is 0.001 (default). The activation function of the hidden layer is tansig and the output node uses the linear transfer function

5 Conclusions

Designing an appropriate model to forecast financial data is a major challenge for time series analysts and researchers. This is mainly because the irregular movements and several changing turning points of these series are practically too difficult to understand and predict. In this study, a new hybrid model that intelligently combines the EMD, PSR and ELM models (EMD + PSR + ELM), is proposed to forecast exchange rates. From the experimental results of this study, we can draw the following conclusions:

(1) EMD can fully capture the local fluctuations of data and can be used as a pre-processor to decompose the complicated raw data into a finite set of IMFs and a residue, which can improve rate predicting accuracy.
(2) The network topology of the model has a major influence on prediction performance for ELM. It is more objective in identifying the chaotic characteristics of exchange rate time series and determining the embedding dimension of the reconstructed phase space through the FNN function. The determined embedding dimension can then be served as the numbers of nodes in the input layer for the SLFN.
(3) Empirical results from four real-world exchange rate time series clearly suggest that our hybrid method substantially improves the overall accuracy of forecasting and also outperformed both a statistical model (Naïve RW) and an artificial intelligence model (ELM). Therefore, the proposed method is extremely suitable for prediction using nonlinear, nonstationary, and highly complex data and is an efficient method for exchange rate prediction.

Future research should consider the property of data in order to combine time series and AI method. Direction prediction criteria are crucial to the trading strategies of investors. In our model, we select only one-dimensional time series of exchange rates for input variables. Future research might attempt to enhance the performance of prediction models by including other efficient input variables such as macroeconomic variables and using diverse data for feasibility. One possibility might try to find important input variables by adopting some strong or emerging mathematical methods, such as MARS, CMARS for building more perfect integrated model. In addition,

the relationships between and trading information about different markets might be examined.

References

Adhikari, R., & Agrawal, R. K. (2013). A combination of artificial neural network and random walk models for financial time series forecasting. *Neural Computing and Applications*, *24*(6), 1441–1449.

Babovic, V., Keijzer, M., & Bundzel, M. (2000). From global to local modelling: A case study in error correction of deterministic models. In *Proceedings of Fourth International Conference on Hydroinformatics*, Iowa City, USA: CD-ROM, IAHR

Bao, Y. K., Xiong, T., & Hu, Z. Y. (2012). Forecasting air passenger traffic by support vector machines with ensemble empirical mode decomposition and slope-based method. *Discrete Dynamics in Nature and Society, 2012*

Barkoulas, J., & Travos, N. (1998). Chaos in an emerging capital market? The case of Athens stock exchange. *Applied Financial Economics*, *8*, 231–243.

Chen, C. F., Lai, M. C., & Yeh, C. C. (2012). Forecasting tourism demand based on empirical mode decomposition and neural network. *Knowledge-Based Systems*, *26*, 281–287. doi:10.1016/j.knosys.2011.09.002.

Chen, F. L., & Ou, T. Y. (2011). Sales forecasting system based on Gray extreme learning machine with Taguchi method in retail industry. *Expert Systems with Applications*, *38*(3), 1336–1345. doi:10.1016/j.eswa.2010.07.014.

Chen, K. L., Yeh, C. C., & Lu, T. L. (2012). A hybrid demand forecasting model based on empirical mode decomposition and neural network in TFT-LCD industry. *Cybernetics and Systems*, *43*(5), 426–441.

Deng, S. K., Yoshiyama, K., Mitsubuchi, T., & Sakurai, A. (2015). Hybrid method mf Multiple kernel learning and genetic algorithm for forecasting short-term foreign exchange rates. *Computational Economics*, *45*(1), 49–89. doi:10.1007/s10614-013-9407-6.

GEP, B., & GM, J. (1970). *Time Series Analysis Forecasting and Control* (3rd ed.). CA: Holden-Day.

Gimore, C. G. (2001). An examination of nonlinear dependence in exchange rates, using recent methods from chaos theory. *Global Finance Journal*, *12*, 139–151.

Hadavandi, E., Shavandi, H., & Ghanbari, A. (2010). Integration of genetic fuzzy systems and artificial neural networks for stock price forecasting. *Knowledge-Based System*, *23*(8), 800–808.

Huang, G. B., Chen, L., & Siew, C. K. (2006a). Universal approximation using incremental constructive feedforward networks with random hidden nodes. *IEEE Transactions on Neural Networks*, *17*(4), 879–892. doi:10.1109/TNN.2006.875977.

Huang, G. B., Zhu, Q. Y., & Siew, C. K. (2006b). Extreme learning machine: Theory and applications. *Neurocomputing*, *70*(1–3), 489–501. doi:10.1016/j.neucom.2005.12.126.

Huang, N. E., Shen, Z., Long, S. R., Wu, M. C., Shih, H. H., Zheng, Q., et al. (1998). The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. *Proceedings of the Royal Society A*, *454*(1971), 903–995.

Huang, N. E., Wu, M. L. C., Long, S. R., Shen, S. S. P., Qu, W., Gloersen, P., et al. (2003). A confidence limit for the empirical mode decomposition and Hilbert spectral analysis. *Proceedings of the Royal Society A*, *459*(2037), 2317–2345.

Huang, S. C., Chuang, P. J., Wu, C. F., & Lai, H. J. (2010). Chaos-based support vector regressions for exchange rate forecasting. *Expert Systems with Applications*, *37*(12), 8590–8598. doi:10.1016/j.eswa.2010.06.001.

Jaeger, H., & Haas, H. (2004). Harnessing nonlinearity: Predicting chaotic systems and saving energy in wireless communication. *Science*, *304*(5667), 78–80.

Kazem, A., Sharifi, E., Hussain, F. K., Saberi, M., & Hussain, O. K. (2013). Support vector regression with chaos-based firefly algorithm for stock market price forecasting. *Applied Soft Somputing*, *13*(2), 947–958. doi:10.1016/j.asoc.2012.09.024.

Kennel, M. B., Brown, R., & Abarbanel, H. D. I. (1992). Determining embedding dimension for phase-space reconstruction using a geometrical construction. *Physical Review A*, *45*(6), 3403–3411.

Lemke, C., & Gabrys, B. (2010). Meta-learning for time series forecasting and forecast combination. *Neurocomputing*, *73*(10–12), 2006–2016.

Liong, S. Y., & Sivapragasam, C. (2002). Flood stage forecasting with SVM. *Journal of the American Water Resources Association*, *38*(1), 173–186.

Liu, H., & Wang, J. (2011). Integrating independent component analysis and principal component analysis with neural network to predict chinese stock market. *Mathematical Problems in Engineering*, *2011*, 15.

Lu, C. J. (2010). Integrating independent component analysis-based denoising scheme with neural network for stock price prediction. *Expert Systems with Applications*, *37*(10), 7056–7064.

Lu, C. J., Lee, T. S., & Chiu, C. C. (2009). Financial time series forecasting using independent component analysis and support vector regression. *Decision Support Systems*, *47*(2), 115–125. doi:10.1016/j.dss.2009.02.001.

Lu, C. J., & Shao, Y. E. (2012). Forecasting computer products sales by integrating ensemble empirical mode decomposition and extreme learning machine. *Mathematical Problems in Engineering*, *2012*, 15.

Makridakis, S. (1993). Accuracy measures: Theoretical and practical concerns. *International Journal of Forecasting*, *9*(4), 527–529.

McKenzie, M. D. (2001). Chaotic behavior in national stock market indices: New evidence from the close returns test. *Global Finance Journal*, *12*, 35–53.

Meese, R., & Rogoff, K. (1983). Empirical exchange rate models of the seventies: Do they fit out of sample? *Journal of International Economics*, *14*(1–2), 3–24.

Rao, C. R., & Mitra, S. K. (1971). *Generalized inverse of matrices and its applications*. New York: Wiley.

Sauer, T., Yorke, J. A., & Casdagli, M. (1991). Embedology. *Journal of Statistical Physics*, *65*(3/4), 579–616.

Sun, Y. (2005). *Exchange rate forecasting with an artificial neural network model: Can we beat a random walk model?*. New Zealand: Lincoln University.

Takens, F. (1981). Detecting strange attractors in turbulence. *Dynamical Systems and Turbulence*, *898*, 366–381.

Tyree, A., & Long, J. (1995). Forecasting currency exchange rates: neural networks and the random walk model. In *Proceedings of the Third International Conference on Artificial Intelligence Applications*. New York: Wall Street

Vasilakis, G. A., Theofilatos, K. A., Georgopoulos, E. F., Karathanasopoulos, A., & Likothanassis, S. D. (2013). A genetic programming approach for EUR/USD exchange rate forecasting and trading. *Computational Economics*, *42*(4), 415–431. doi:10.1007/s10614-012-9345-8.

Wang, J. D., & Qi, W. G. (2009). Prediction of river water turbidity based on EMD-SVM. *Acta Electronica Sinica*, *37*(10), 2130–2133.

Wang, J. J., Wang, J. Z., Zhang, Z. G., & Guo, S. P. (2012). Stock index forecasting based on a hybrid model. *Omega*, *40*(6), 758–766. doi:10.1016/j.omega.2011.07.008.

Wu, J. L., & Chang, P. C. (2012). A trend-based segmentation method and the support vector regression for financial time series forecasting. *Mathematical Problems in Engineering*, *2012*, 1–20.

Xia, M., Zhang, Y., Weng, L., & Ye, X. (2012). Fashion retailing forecasting based on extreme learning machine with adaptive metrics of inputs. *Knowledge-Based Systems*, *36*, 253–259. doi:10.1016/j.knosys.2012.07.002.

Xuan, Z. Y., & Yang, G. X. (2008). Application of EMD in the atmosphere time series prediction. *Acta Automat Sinica*, *34*(1), 97–101.

Yang, Y. F., Bao, Y. K., Hu, Z. Y., & Zhang, R. (2010). Crude oil price prediction based on empirical mode decomposition and support vector machines. *Chinese Journal of Management*, *7*(12), 1884–1889.

Ye, L., & Liu, P. (2011). Combined model based on EMD-SVM for short-term wind power prediction. *Proceedings of the Chinese Society of Electrical Engineering*, *31*(31), 102–108.

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

Zhang, Y., & Wu, L. (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. doi:10.1016/j.eswa.2008.11.028.

Zhiqiang, G., Huaiqing, W., & Quan, L. (2012). Financial time series forecasting using LPP and SVM optimized by PSO. *Soft Computing*, *17*(5), 805–818.

Zhu, B. Z., & Wei, Y. M. (2013). Carbon price forecasting with a novel hybrid ARIMA and least squares support vector machines methodology. *Omega*, *41*, 517–524.