Received: 26 December 2019 | Revised: 21 June 2020 | Accepted: 23 July 2020
DOI: 10.1111/exsy.12633

ORIGINAL ARTICLE

Expert Systems WILEY

A hybrid model for financial time-series forecasting based on mixed methodologies

Zhidan Luo[1] | Wei Guo[1] | Qingfu Liu[2] | Zhengjun Zhang[3]

[1]School of Statistics, University of International Business and Economics, Beijing, China
[2]Institute for Financial Studies, Fudan University, Shanghai, China
[3]Department of Statistics, University of Wisconsin-Madison, Madison, Wisconsin

Correspondence
Qingfu Liu, Institute for Financial Studies, Fudan University, Shanghai, China.
Email: liuqf@fudan.edu.cn

Funding information
National Natural Science Foundation of China, Grant/Award Numbers: 61973084, 71871066, 71991471; NSF-DMS, Grant/Award Number: 2012298; the Shanghai Science and Technology Innovation Action Plan Project, Grant/Award Number: 19511101700

Abstract

This paper proposes a hybrid model that combines ensemble empirical mode decomposition (EEMD), autoregressive integrated moving average (ARIMA), and Taylor expansion using a tracking differentiator to forecast financial time series. Specifically, the financial time series is decomposed by EEMD into some subseries. Then, the linear portion of each subseries is forecasted by the linear ARIMA model, while the nonlinear portion is predicted by the nonlinear Taylor expansion model. The forecasting results of the linear and nonlinear models are combined into the predicted result of each subseries. The final prediction result is obtained by combining the prediction values of all the subseries. The empirical results with real financial time series data demonstrate that this new hybrid approach outperforms the benchmark hybrid models considered in this paper.

KEYWORDS
ARIMA, EEMD, financial time series, forecasting, Taylor expansion

1 | INTRODUCTION

Financial time series forecasting is a challenging task. It has attracted considerable attention in academic and practical fields (Ahmad & Mohd, 2017). As a particular kind of time series, financial time series are intrinsically complex, dynamic, noisy, and nonstationary, which make them difficult to forecast. The existing literature has many studies on time series forecasting with different models. In general, these models can be divided into two categories (Wang, Wang, Zhang, & Guo, 2012). The first is traditional prediction models based on statistical principles, such as autoregressive integrated moving average (ARIMA) (Box & Jenkins, 1970) and generalized autoregressive conditional heteroskedasticity (Bollerslev, 1986). The second includes the new methods developed on the basis of artificial intelligence (AI). AI has had an increasingly wide application in a wide range of research fields (Lu, 2019), where artificial neural networks (ANN) and support vector machines (SVM) are currently popular AI algorithms which can be used in classification and forecasting (Akman, Karaman, & Kuzey, 2020; Crone, Hibon, & Nikolopoulos, 2011; Kim, 2003; Kock & Teräsvirta, 2014; Pai & Lin, 2005; Shi, Xu, & Liu, 1996; Shi, Xu, & Liu, 1999; Vafeiadis et al., 2018; Wegener, Spreckelsen, Basse, & Mettenheim, 2016; Xie, Mao, & Wang, 2015; Yuan, Li, Guan, & Xu, 2010; Zhang, 2003).

A single linear or nonlinear model is not always adequate to determine all characteristics of a time series. Some efforts have been made to develop hybrid models to overcome the shortcoming. An earlier effort to establish a hybrid approach using both ARIMA and ANN models to predict time series is made in Zhang (2003). First, the linear structure is extracted with an ARIMA model, while an ANN model predicts the nonlinear component. Then, the two predicted results are combined as the final predicted result. The results with real data sets indicate that the hybrid model can effectively improve forecasting accuracy. However, in practice, the neural network algorithm suffered some problems with over-fitting and local minimums. Based on the structured risk minimization principle, SVM is a kind of machine learning method, which often has good prediction results in nonlinear time series forecasting. Pai and Lin (2005) develop a hybrid model combining ARIMA and SVM, which takes a full advantage of the unique strengths of the two models. They use real data sets of stock prices to examine the forecasting accuracy of the proposed model and find that the experimental results are very promising. Although SVM can address the problems of small samples, nonlinearity, over-fitting, the curse of dimensionality, and local minimums, SVM is sensitive to kernel functions and does not consider the correlation among sequences in time series, which limits its applications in financial time series.

A hybrid model using ARIMA and Taylor expansion forecasting (TEF) based on a tracking differentiator (TD) is proposed to predict commodity's future prices in Zhang, Zhang, and Feng (2016). TD is a controller design method in control theory, first developed by Han and Wang (1994) to extract derivative information on signals with noise interference. An earlier attempt to use a TEF model based on TD to predict the stock price was in Liu, Wang, and Zhang (2006). The empirical results in Zhang et al. (2016) show that the ARIMA-TEF model achieves better predicted results than the ARIMA-SVM model.

Financial time series show the characteristics of nonlinearity, nonstationarity, and being multiscale, which pose great challenges in forecasting them. However, decomposing a complex time series into a set of simple modes with multiscale, simple, stationary, and regular characteristics using empirical mode decomposition (EMD) simplifies the forecasting task and is suitable for forecasting time series. Introduced by Huang et al. (1998), EMD is an empirical, intuitive, direct, and self-adaptive data processing method. It not only works with nonlinear and nonstationary data but also provides economic meaning. Ensemble empirical mode decomposition (EEMD) is an improved EMD (Wu & Huang, 2009), which addresses the problem of mode mixing in EMD. The models with EMD or EEMD have been widely applied to time series analysis and prediction (Kožić & Sever, 2014; Plakandaras, Gupta, Gogas, & Papadimitriou, 2015; Tiwari, Dar, Bhanja, & Gupta, 2016; Wang, Hu, Wu, Liu, & Bai, 2014; Wang, Wang, Song, & Liu, 2018; Yang & Lin, 2016; Yang & Lin, 2017; Zhang, Lin, & Shang, 2017; Zhang, Wei, Tan, Wang, & Tian, 2017; Zhu, Shi, Chevallier, Wang, & Wei, 2016). Zhang, Lai, and Wang (2008) analyze crude oil prices using the EEMD method and show that the EEMD is a vital technique for analyzing crude oil prices. Zeng, Qu, Ng, and Zhao (2016) develop an approach based on EMD and ANN for forecasting the Baltic Dry Index. The prediction results demonstrate that the proposed EMD-ANN method outperforms ANN and vector autoregression. Zhang, Wei, et al. (2017) propose a hybrid model combining EEMD, an adaptive neural network-based fuzzy inference system, and a seasonal autoregressive integrated moving average for short-term forecasting of wind speeds. Their numerical testing results based on two wind sites in South Dakota shed light on the effectiveness of the hybrid method. Wang et al. (2018) use an EEMD-based model to forecast coal overcapacity, and their empirical results indicate that the proposed model significantly outperforms other widely developed baselines. Yang and Lin (2017) propose a new hybrid model that intelligently combines the EMD, phase space reconstruction (PSR), and extreme learning machine (ELM) models (EMD-PSR-ELM) to forecast exchange rates. Their experimental results with real-world exchange rate time series show that the proposed approach yields superior results than the benchmark prediction models, including Naïve random-walk, ARIMA, back-propagation neural network (BPNN), ELM, EMD-ELM, and PSR-ELM models.

Although the above-discussed hybrid models exhibit much better forecasting performance in various applications, more accurate prediction models are still in demand considering the increasing requirements for better prediction of financial time series. As a special kind of data, financial time series forecasting is regarded as one of the most challenging jobs in forecasting owing to its inherent complexity. To improve the accuracy of the forecast model, this study is to develop a hybrid model for predicting financial time series. On one hand, it has been known the ARIMA model is one of the most popular and widely used time series models due to its favourable statistical properties. On the other hand, compared with the SVM, the TEF model can improve the forecast accuracy. Given that the ARIMA model cannot capture nonlinear patterns, and nonlinear models, including the TEF model, are not adequate in modelling and forecasting linear time series. A natural idea is to integrate them in a new model. In addition, the EEMD method has been known for its strong performance in prediction. As such, it should be considered before building prediction models. These observations led us to propose a new hybrid forecasting model which possesses enhanced prediction capability to forecast financial time series by combining EEMD, TEF, and ARIMA in this paper.

The main reasons for combining EEMD, ARIMA, and TEF can be summarized as follows. First, using hybrid models or combining several models has become a common practice to improve the forecasting accuracy where each method's unique strengths are combined to capture different characteristics in the data. Second, the EEMD can reveal the hidden patterns in time series, and it tends to assist in designing forecasting models. It is natural to include the EEMD-based model for forecasting time series. Third, the ARIMA model has achieved successes in linear domains; so far, it is still a widely used linear model. Fourth, the TEF model owns good strengths, but there are fewer researches on hybrid models based on the TEF model. It is worth to make further efforts to study the TEF-based hybrid models and extend their applications.

We first use EEMD to decompose financial time series into some subseries with independent intrinsic modes, including some intrinsic mode functions (IMFs) and one residual. Then, we forecast the linear component of each subseries with the ARIMA model, while the nonlinear component is predicted with the TEF model. The forecasting results of the two models are combined as the predicted results of the subseries. Ultimately, the predicted results of all the subseries are combined to obtain the final prediction result of the original financial time series.

In summary, the main contributions of this paper are as follow: (a) a novel hybrid model combining EEMD, ARIMA and TEF model is developed to forecast the financial time series, and the model developed by Zhang et al. (2016) and the novel EMD-PSR-ELM model proposed by Yang and Lin (2017) are selected as benchmark models because of their validity and strong performance; (b) to illustrate the extensive applicability, we use the proposed hybrid model to forecast real-world financial time series, for example, stock indexes which are widely used in the world; (c) besides the three error criteria commonly used in Zhang et al. (2016) and Yang and Lin (2017), a statistical measure on directional prediction accuracy is also used to evaluate the predictive accuracy of different models; (d) we conduct experiments with different testing data sets to show the superior performance and robustness of the new hybrid EEMD-ARIMA-TEF model; (e) an artificial investment strategy called long-short strategy is considered to show the usefulness of the forecast by the proposed EEMD-ARIMA-TEF model.

The rest of this paper is organized as follows: Section 2 introduces the basic principles of EEMD, ARIMA, and TEF models based on TD respectively. Section 3 explicates the design of the new hybrid model. Section 4 presents the empirical results, and Section 5 concludes.

2 | METHODOLOGY

2.1 | EEMD

EMD, a self-adaptive decomposition technique, is effective in extracting the characteristic information from nonstationary and nonlinear time series, for example, financial time series. This technique has several evident advantages. First and foremost, EMD can decompose any nonstationary and nonlinear data into simple independent IMFs. Second, this decomposition technology is based on local characteristic time scales in time series data, and only the extrema are extracted in the sifting process; therefore, EMD is local, self-adaptive, concretely implicational, and highly efficient. For comparison, alternatively, wavelet-based models can be used for time series forecasting. For example, Huang and Wu (2008) propose a wavelet-based hybrid model to forecast stock indexes. It is known that wavelet decomposition techniques need to determine a filter function before decomposition. However, EMD is not flawless. The main drawback of the EMD method is the mode-mixing problem. To address this problem, the EEMD method is proposed by Wu and Huang (2009) as follows:

(1) Initialize the number of ensemble M and the standard deviation of added white noise ε, set $i = 1$.

(2) Add the white noise series to the original financial time series $\{y(t)\}_{t=1}^{N}$, and obtain a new series $\{y_i(t)\}_{t=1}^{N}$:

$$y_i(t) = y(t) + n_i(t) \tag{1}$$

where $\{n_i(t)\}_{t=1}^{N}$ denotes the added white noise series i, and $\{y_i(t)\}_{t=1}^{N}$ represents the noise-added series i.

(3) Identify all the local maxima and minima of time series $\{y_i(t)\}_{t=1}^{N}$, and generate the upper envelopes and lower envelopes of $\{y_i(t)\}_{t=1}^{N}$ with cubic spline interpolation.

(4) Calculate the point-by-point mean from upper and lower envelopes and record as $m(t)$, and then extract the difference between $y_i(t)$ and $m(t)$ as $h(t)$:

$$h(t) = y_i(t) - m(t) \tag{2}$$

(5) Check the properties of $h(t)$. If $h(t)$ is an IMF, it meets the following two conditions:
  (a) The number of extreme values and the number of zero crossings either are equal or differ at most by one;
  (b) The mean value of the envelope defined by the local maxima and the envelope defined by the local minima is zero at any point. Take $h(t)$ as the first IMF c_{i1} and replace $y_i(t)$ with the residual $r_i(t) = y_i(t) - c_{i1}$. If $h(t)$ is not an IMF, replace $y_i(t)$ with $h(t)$.

(6) Repeat Steps (3)–(5) until the residual satisfies the stopping criterion and obtain all IMFs c_{ij} (j = 1, 2, …, n, n is the number of IMFs) and a residual $r_i(t)$.

(7) Set $i = i + 1$ and repeat Steps (2)–(6) with different white noise series until $i = M$.

(8) Calculate the combined mean of the M trials for the corresponding IMFs and residual obtained from each decomposition, and take the mean as the final decomposition results.

2.2 | ARIMA

The ARIMA model, introduced by Box and Jenkins, is one of the most popular approaches to time series forecasting. In an ARIMA model, the future value of the time series $\{y_t\}_{t=1}^{N}$ is a linear combination of past values and past errors and it can be expressed as follows:

$$y_t = \theta_0 + \varphi_1 y_{t-1} + \varphi_2 y_{t-2} + \cdots + \varphi_p y_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \cdots - \theta_q \varepsilon_{t-q} \tag{3}$$

where ε_t is the residual term between the actual data and the forecasting value; φ_i and θ_j (i = 1, 2, …, p, j = 1, 2, …, q) are autoregressive coefficients and moving average coefficients to be estimated. P and q are integers, which are often referred to as orders of the autoregressive and moving average models, respectively. In general, this model has three phases: model identification, parameter estimation, and diagnostic checking. The ARIMA model can be used to predict time series after diagnostic checking is conducted.

2.3 | TEF

The Taylor formula is as follows:

$$f(t_i) = f(t_{i-1}+h) = f(t_{i-1}) + \dot{f}(t_{i-1})h + \frac{1}{2}\ddot{f}(t_{i-1})h^2 + \cdots + \frac{1}{n!}f^{(n)}(t_{i-1})h^n + \frac{1}{(n+1)!}f^{(n+1)}(t_{i-1}+\theta h)h^{n+1},\ \theta \in (0,1). \tag{4}$$

When h is small enough, Equation (4) can be approximated as:

$$f(t_{i-1}+h) \approx f(t_{i-1}) + \dot{f}(t_{i-1})h + \frac{1}{2}\ddot{f}(t_{i-1})h^2 + \cdots + \frac{1}{n!}f^{(n)}(t_{i-1})h^n \tag{5}$$

Here, we assume that h represents the time interval, and t_i represents a moment. If we know the values of $f(t)$ at t_{i-1} and its all-order derivatives, then we can obtain the value of $f(t)$ at $t_{i-1} + h$ using Equation (5), which means that we can forecast the value $f(t_{i-1} + h)$ at time t_{i-1}, with error

$$\frac{1}{(n+1)!}f^{(n+1)}(t_{i-1}+\theta h)h^{n+1} \tag{6}$$

In general, financial time series are neither smooth nor differentiable but can still be predicted with the Taylor expansion model (see Remark 1 in Zhang et al., 2016). It is assumed that any given time series $\{y_i\}_{i=1}^N$ can be sampled by function φ with step size h, that is, $\varphi(0) = y_1$, $\varphi(h) = y_2$, $\cdots$, $\varphi((N-1)h) = y_N$. To forecast y_{N+1} using Equation (5), we need to know $\varphi((N-1)h)$ and its all-order derivatives. Because of the sensitivity to noise, however, the derivative of the usually rapidly varying noise will “drown out” the derivative of the original information and it is not available to compute the derivative directly. Fortunately, the derivatives of φ at $(N-1)h$ can be estimated with the TD (Guo, Han, & Xi, 2002; Guo & Zhao, 2011). Estimating the derivatives with the TD and then using Taylor expansion, one-step-ahead prediction can be achieved, which is called the TEF model based on a TD. Next, a key problem is the construction of the TEF model using a TD. For the sake of comparison, we still apply the following third-order high-gain TD used in Zhang et al. (2016):

$$\begin{cases} \dot{z}_1(t) = z_2(t) - 3r[z_1(t) - \varphi(t)] \\ \dot{z}_2(t) = z_3(t) - 6r^2[z_1(t) - \varphi(t)] \\ \dot{z}_3(t) = -6r^3[z_1(t) - \varphi(t)] \\ z_1(0) = z_{10}, z_2(0) = z_{20}, z_3(0) = z_{30} \end{cases} \tag{7}$$

where z_{10}, z_{20}, and z_{30} are given initial values, $\varphi(t)$ is input, $z_1(t)$, $z_2(t)$, and $z_3(t)$ are output used to approximate $\varphi(t)$, $\dot{\varphi}(t)$, and $\ddot{\varphi}(t)$, respectively, and $r > 0$ is the gain. The corresponding forecasting strategy of Equation (7) can be designed as follows:

$$y_{N+1} = \varphi(Nh) \approx y_N + z_2 h + \frac{1}{2}z_3 h^2. \tag{8}$$

In order to use the forecasting model (8), we discretize system (7) to be

$$\begin{cases} Z_1(i+1) = Z_1(i) + Z_2(i)h - 3r[Z_1(i) - \varphi(i)]h \\ Z_2(i+1) = Z_2(i) + Z_3(i)h - 6r^2[Z_1(i) - \varphi(i)]h \\ Z_3(i+1) = Z_3(i) - 6r^3[Z_1(i) - \varphi(i)]h \\ Z_1(1) = z_{10}, Z_2(1) = z_{20}, Z_3(1) = z_{30} \end{cases} \tag{9}$$

And our forecasting strategy is easily to get:

$$y_{N+1} = y_N + Z_2(N)h + \frac{1}{2}Z_3(N)h^2. \tag{10}$$

3 | DESIGN OF THE NEW HYBRID MODEL

This section is devoted to the design of the proposed hybrid EEMD-ARIMA-TEF model. As discussed earlier, due to the complex features, forecasting financial time series is challenging. Deep insights into the original financial time series are important for achieving more accurate prediction results. Based on a decomposition and combination prediction method, the hybrid EEMD-ARIMA-TEF model with both nonlinear and linear modelling capabilities is a good choice for forecasting financial time series. The procedures are as follows:

(1) The original financial time series $\{y(t)\}_{t=1}^{N}$ is first decomposed into some IMFs and a residual series with the EEMD method:

$$y(t) = \sum_{j=1}^{n} c_j(t) + r(t), \tag{11}$$

where n is the number of IMFs, and $c_j(t)$ and $r(t)$ represent the IMFs and the residual series, respectively.

(2) Following the well-established "linear and nonlinear" modelling framework (Pai & Lin, 2005; Zhang, 2003), each subseries $x_i(t)$ (i.e., $c_j(t)$ and $r(t)$) can be represented as follows:

$$x_i(t) = l(t) + nl(t), \tag{12}$$

where $l(t)$ is the linear part, and $nl(t)$ is the nonlinear part of $x_i(t)$. We use ARIMA to model $l(t)$, and the residual can be obtained from the ARIMA model:

$$e(t) = x_i(t) - \hat{l}(t), \tag{13}$$

where $\hat{l}(t)$ is the forecasted value from ARIMA, and $e(t)$ represents the residual.

(3) The residual $e(t)$ contains only the nonlinear relationship. The nonlinear TEF model for $e(t)$ can be represented as follows:

$$e(t) = f(e(t-1), e(t-2), \ldots, e(1)) + \varepsilon(t), \tag{14}$$

where f is the nonlinear function defined by the TEF model, and $\varepsilon(t)$ is the error term. The nonlinear part $nl(t)$ can be forecasted and remembered as $\hat{nl}(t)$. Then the combined forecast is:

$$\hat{x}_i(t) = \hat{l}(t) + \hat{nl}(t), \tag{15}$$

where $\hat{x}_i(t)$ is the forecasted value of $x_i(t)$.

(4) The final prediction results are obtained by combining the prediction values $\hat{x}_i(t)$:

$$\hat{y}(t) = \sum_{i=1}^{n+1} \hat{x}_i(t), \tag{16}$$

where $\hat{y}(t)$ is the forecasted value of the original financial time series $y(t)$.

The proposed EEMD-ARIMA-TEF method can also be illustrated in the flow chart in Figure 1.

4 | EMPIRICAL RESULTS

4.1 | Data

To evaluate the performance of the proposed forecasting model, we base our analysis on daily closing price data sets of stock indexes—namely, SP500, NIKKEI 225, AORD, and CSI300. The first three data sets are downloaded from Yahoo Finance (http://finance.yahoo.com), and the last data sets are downloaded from NetEase Finance (https://money.163.com/). These entire daily closing price data sets from January 1, 2010, to December 31, 2019, are used and illustrated in Figure 2, which indicates that the stock index prices are highly uncertain, nonlinear, and dynamic. Table 1 presents some brief descriptive statistics for the four stock indexes. These statistics related to skewness, kurtosis, and Jarque-Bera tests reveal that these stock index prices are all non-normal. To verify the effectiveness of the proposed hybrid model, we predict the prices on the last 30 trading days on the four stock indexes based on past data.

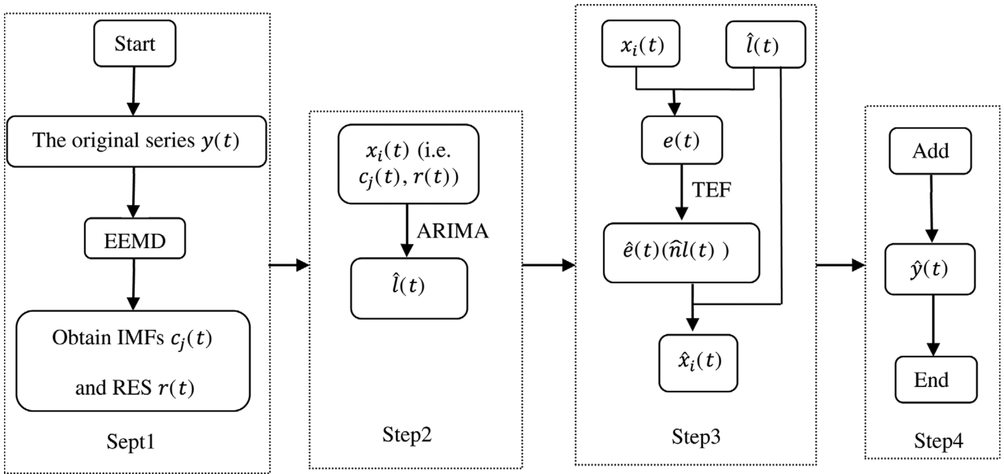


FIGURE 1 The procedures in the proposed EEMD-ARIMA-TEF model. ARIMA, autoregressive integrated moving average; EEMD, ensemble empirical mode decomposition; TEF, Taylor expansion forecasting

4.2 | Performance criteria

To evaluate the forecasting performance of different models, we use the three error measures used in Zhang et al. (2016) and Yang and Lin (2017): mean absolute error (MAE), mean absolute percentage error (MAPE), and root mean squared error (RMSE). The definitions of these error criteria are as follows:

$$\mathrm{MAE} = \frac{1}{n}\sum_{t=1}^{n}|y_t - \hat{y}_t|, \tag{17}$$

$$\mathrm{MAPE} = \frac{100}{n}\sum_{t=1}^{n}\left|\frac{y_t - \hat{y}_t}{y_t}\right|, \tag{18}$$

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{t=1}^{n}(y_t - \hat{y}_t)^2}. \tag{19}$$

where y_t and $\hat{y}_t$ are the actual value and predicted value, respectively, and n is the number of prediction samples. Obviously, MAE, MAPE, and RMSE measure the deviation between actual values and predicted values. Hence, the model's forecasting performance is better when these measures are lower.

In addition, it is important to predict the future trends of the stock indexes correctly, especially for investors. Therefore, it is still needed to judge whether the prediction model can predict the trends well by performance criteria. DS (directional symmetry) is a statistical measure on trend, which measures the directional prediction accuracy. DS can be expressed as:

$$DS = \frac{1}{n}\sum_{t=1}^{n}d_t, \quad \text{where } d_t = \begin{cases} 1, & (\hat{y}_t - \hat{y}_{t-1})(y_t - y_{t-1}) \geq 0. \\ 0, & \textit{otherwise}. \end{cases} \tag{20}$$

It is clearly that a bigger value of DS suggests a better predictor.

4.3 | Implementation of models

As mentioned in Section 1, this study adopts the ARIMA-TEF and EMD-PSR-ELM as benchmarks in our experiment. The use of these two benchmarks is detailed in Zhang et al. (2016) and Yang and Lin (2017).

Then, we employ our EEMD-ARIMA-TEF model, illustrated in Section 3. It is noteworthy that *M* (i.e., the ensemble number) and ε (i.e., the standard deviation of added white noise) are set as 100 and 0.1, respectively, as shown in Wu and Huang (2009). The decomposition results are in Figures 3–4. The IMFs are sorted by fluctuation frequency from high to low. There is almost no fluctuation in the residual series, as shown at the bottom of Figures 3–4, which means that the long-term trends in the closing prices remain relatively stable without interference factors.

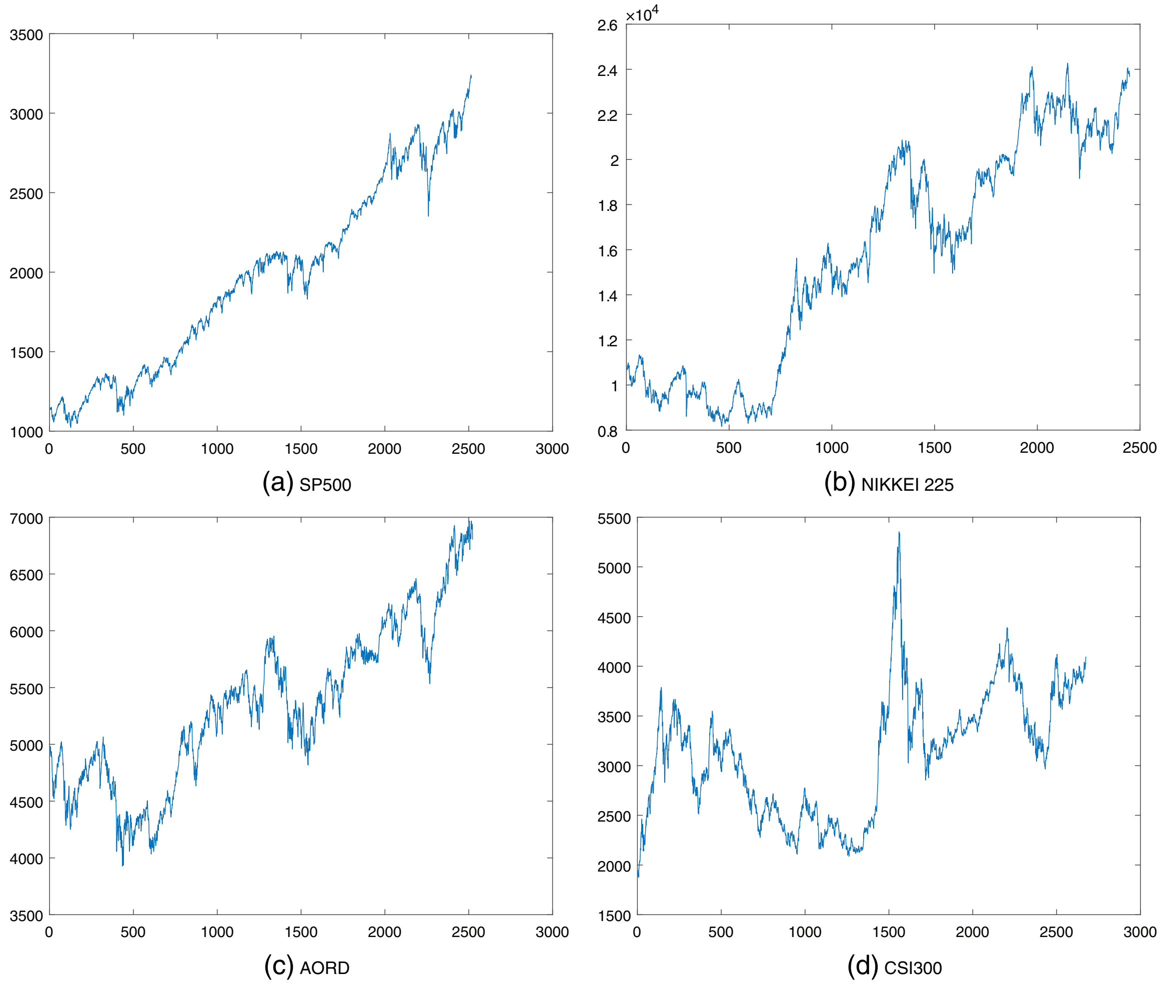


FIGURE 2 Daily stock indexes from 2010 to 2019

TABLE 1 Descriptive statistics of the stock index data

Stock indexes	SP500	NIKKEI 225	AORD	CSI300
Mean	1962.609	15,786.66	5,341.172	3,129.433
Median	1986.480	16,178.94	5,351.500	3,188.202
Maximum	3,240.020	24,270. 62	6,967.000	5,353.751
Minimum	1,022.580	8,160.010	3,927.600	1876.185
Std. dev.	588.9103	4,933.657	700.2314	627.8903
Skewness	0.203017	−0.096734	0.184720	0.267856
Kurtosis	1.897014	1.592367	2.327263	2.682760
Jarque-Bera	144.8212	206.0078	61.92499	43.20454
p-value	0.000000	0.000000	0.000000	0.000000
Observations	2,516	2,449	2,523	2,675

4.4 | Prediction results

This section is devoted to predicting the closing price on the four stock indexes (SP500, NIKKEI 225, AORD, and CSI300) using the proposed EEMD-ARIMA-TEF method and the two benchmark models.

Our study considers only one-step-ahead forecasting, which avoids the problems associated with cumulative errors from previous forecast periods. The forecasting results on the four stock indexes are reported in the following analysis.

The actual values and the predicted results of the four stock indexes by the three hybrid models with the testing data within 30 days are shown in Figure 5. The point-to-point comparisons of the experimental results in Figure 5 show that the predicted values of the

EEMD-ARIMA-TEF model are closer to the real data than those of the ARIMA-TEF and EMD-PSR-ELM models, especially for some local extreme values. Also, the predicted results of the ARIMA-TEF model deviate from the original values severely. In contrast with the other two benchmarks, our EEMD-ARIMA-TEF model relatively performs well according to the graph.

To compare the performance of different models further, the three error measures (MAE, MAPE, and RMSE) and DS are calculated and reported in Table 2. We first compare our EEMD-ARIMA-TEF model with the ARIMA-TEF model. It is evidently found that all the testing MAE in EEMD-ARIMA-TEF are much smaller than those in ARIMA-TEF, for example, 6.9881 and 13.7720 in SP500, respectively. The testing MAPE and RMSE are also much smaller in our EEMD-ARIMA-TEF model than the ARIMA-TEF model. These results indicate that the deviation between the actual and predicted values using our EEMD-ARIMA-TEF model is smaller. On the contrary, all testing DS in EEMD-ARIMA-TEF in SP500, NIKKEI 225, AORD, and CSI300 are much larger than those in ARIMA-TEF, which means the proposed EEMD-ARIMA-TEF model can provide the consistency in the prediction of stock index trend. It can be seen that the impact of EEMD on ARIMA-TEF is significant, which shows once again, the EEMD step has improved the forecasting

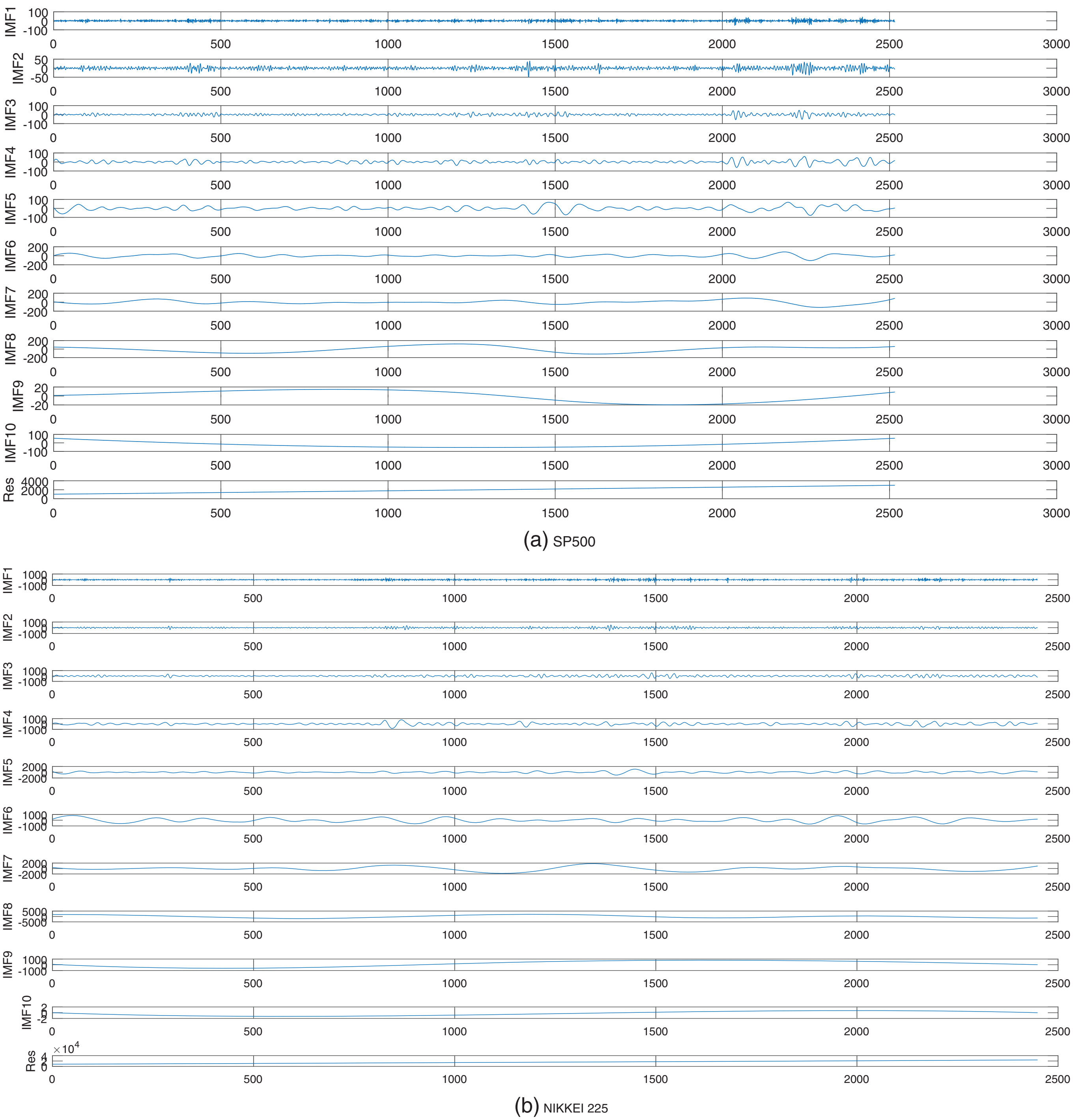


FIGURE 3 Decomposition results. Note: M = 100, ε = 0.01

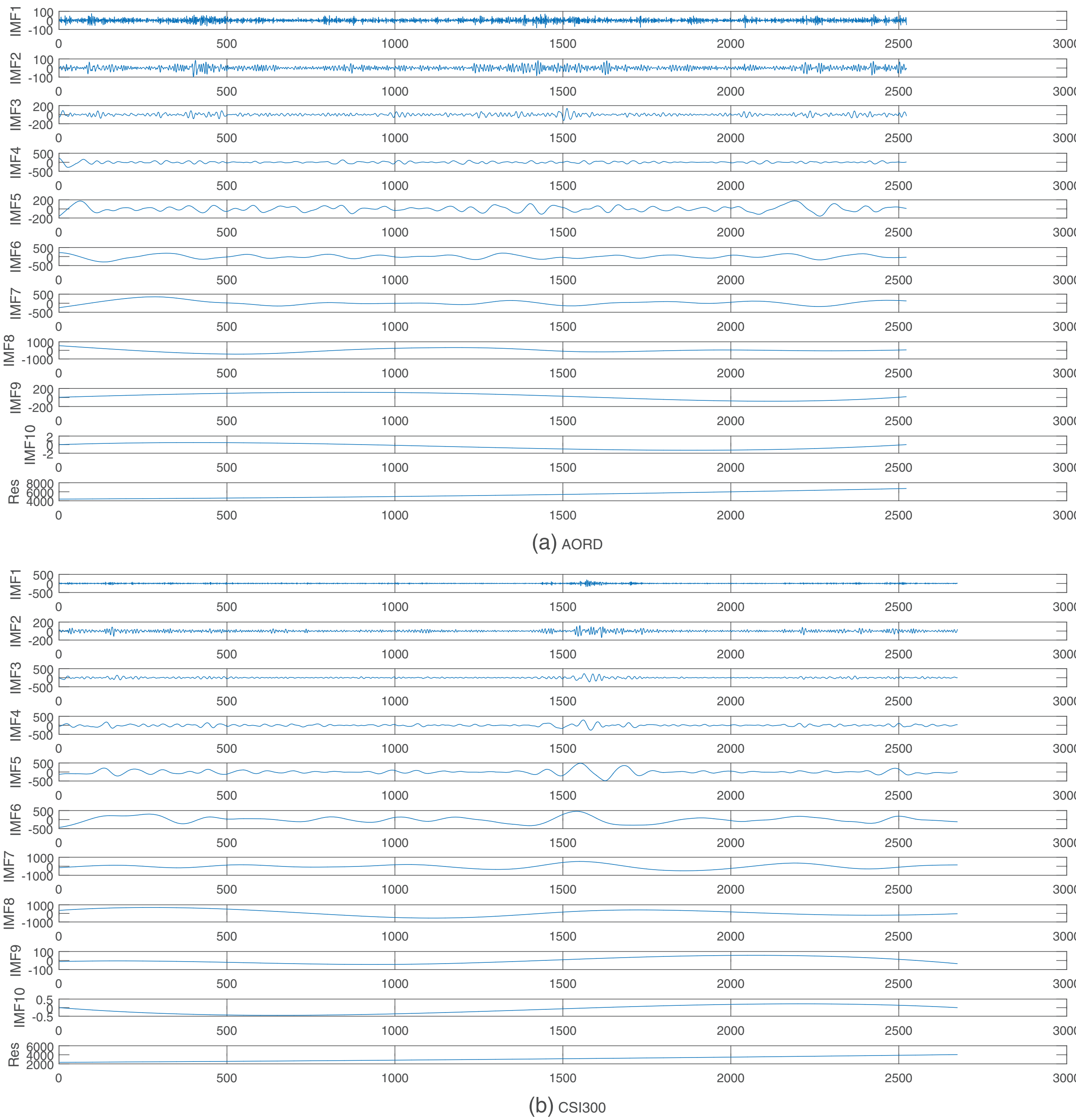


FIGURE 4 Decomposition results. Note: M = 100, ε = 0.01

performance. Next, we can see that all the testing MAE in the EMD-PSR-ELM model are 9.6214 (SP500), 94.3270 (NIKKEI 225), 27.6464 (AORD), and 16.6772 (CSI300), whereas those of our EEMD-ARIMA-TEF model are much lower, at 6.9881, 52.7123, 17.8508, and 10.8782, respectively. In addition to MAE, the testing MAPE and RMSE are also much smaller using our model. Moreover, our model has higher testing DS, that is, 0.8621 (SP500), 0.7931 (NIKKEI 225), 0.7931 (AORD), and 0.7931 (CSI300), while the DS in EMD-PSR-ELM are 0.6207 (SP500), 0.6207 (NIKKEI 225), 0.7241 (AORD), and 0.5172 (CSI300), respectively. Therefore, it can be concluded that our model can provide a smaller deviation than EMD-PSR-ELM and better in direction prediction. These results of all the testing performance criteria demonstrate that our proposed model boasts much more excellent prediction performance than the ARIMA-TEF and EMD-PSR-ELM models.

In short, both the point-to-point comparisons and all criteria indicate that our EEMD-ARIMA-TEF model results in better prediction performance than the ARIMA-TEF and EMD-PSR-ELM models.

To show our model is robust, we further predict the prices on the last 60, 100, and 200 trading days of SP500, NIKKEI 225, AORD, and CSI300 stock indexes. The prediction results are in Figure 6 and Table 2. We observe again that our EEMD-ARIMA-TEF model outperforms the

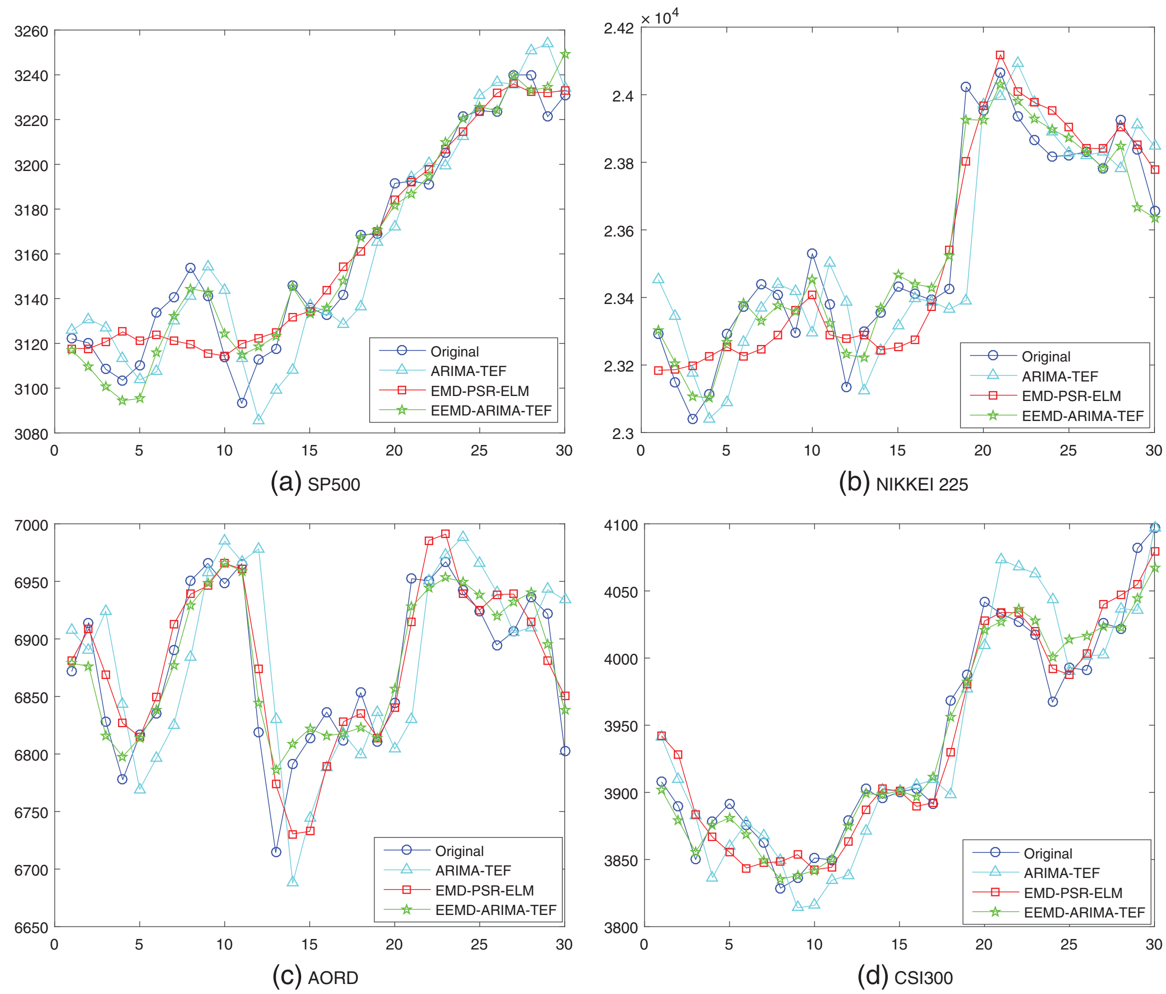


FIGURE 5 Comparison of the predicted results of closing price

TABLE 2 Performance comparison

Forecast period (no. of days)	Stock index	Performance criteria	ARIMA-TEF	EEMD-ARIMA-TEF	EMD-PSR-ELM
30	SP500	MAE	13.7720	6.9881	9.6214
		MAPE	0.4367	0.2219	0.3059
		RMSE	17.1044	8.9580	12.6945
		DS	0.5862	0.8621	0.6207
	NIKKEI 225	MAE	123.8213	52.7123	94.3270
		MAPE	0.5258	0.2237	0.4013
		RMSE	169.8337	64.8000	110.2595
		DS	0.5172	0.7931	0.6207
	AORD	MAE	51.6293	17.8508	27.6464
		MAPE	0.7540	0.2603	0.4035
		RMSE	65.8928	22.6568	34.6473
		DS	0.4828	0.7931	0.7241
	CSI300	MAE	25.7858	10.8782	16.6772
		MAPE	0.6535	0.2738	0.4244
		RMSE	32.1403	14.7213	20.2670
		DS	0.5517	0.7931	0.5172

TABLE 2 (Continued)

Forecast period (no. of days)	Stock index	Performance criteria	ARIMA-TEF	EEMD-ARIMA-TEF	EMD-PSR-ELM
60	SP500	MAE	14.6077	6.2591	9.3901
		MAPE	0.4757	0.2029	0.3049
		RMSE	18.7060	8.2143	12.0909
		DS	0.5424	0.7797	0.6102
	NIKKEI 225	MAE	129.1930	61.6058	161.9364
		MAPE	0.5623	0.2678	0.6963
		RMSE	177.5335	76.4401	193.4958
		DS	0.5254	0.8305	0.6949
	AORD	MAE	45.8208	15.5201	27.3894
		MAPE	0.6721	0.2273	0.4017
		RMSE	58.7597	19.2036	33.7882
		DS	0.5254	0.8136	0.6610
	CSI300	MAE	28.2038	10.0629	16.4636
		MAPE	0.7177	0.2549	0.4198
		RMSE	34.9175	13.2739	20.1328
		DS	0.5254	0.8136	0.5763
100	SP500	MAE	19.0311	7.2592	11.4511
		MAPE	0.6365	0.2415	0.3829
		RMSE	25.1806	9.2996	15.8441
		DS	0.4848	0.8182	0.5758
	NIKKEI 225	MAE	139.9219	62.4513	421.0170
		MAPE	0.6339	0.2826	1.8358
		RMSE	182.4020	77.4607	499.3559
		DS	0.4949	0.8283	0.6768
	AORD	MAE	45.4006	15.1282	26.8498
		MAPE	0.6721	0.2236	0.3977
		RMSE	58.7232	19.2352	33.8894
		DS	0.4949	0.8182	0.6465
	CSI300	MAE	28.6919	10.8809	17.6248
		MAPE	0.7395	0.2801	0.4551
		RMSE	36.1295	14.4047	21.3945
		DS	0.4747	0.8485	0.5758
200	SP500	MAE	18.9567	7.8950	11.9432
		MAPE	0.6450	0.2687	0.4057
		RMSE	25.2779	9.9751	15.8486
		DS	0.5176	0.8141	0.6382
	NIKKEI 225	MAE	147.9214	66.4041	217.5344
		MAPE	0.6820	0.3060	0.9665
		RMSE	195.7740	84.4125	290.1773
		DS	0.4975	0.8291	0.6533
	AORD	MAE	42.8100	14.5655	25.0036
		MAPE	0.6430	0.2190	0.3744
		RMSE	54.6670	18.7123	31.6635
		DS	0.5276	0.8543	0.6683
	CSI300	MAE	37.7335	14.4691	22.5798
		MAPE	0.9851	0.3773	0.5894
		RMSE	50.5256	19.1782	29.1263
		DS	0.4925	0.8291	0.6281

Abbreviations: ARIMA, autoregressive integrated moving average; EEMD, ensemble empirical mode decomposition; TEF, Taylor expansion forecasting.

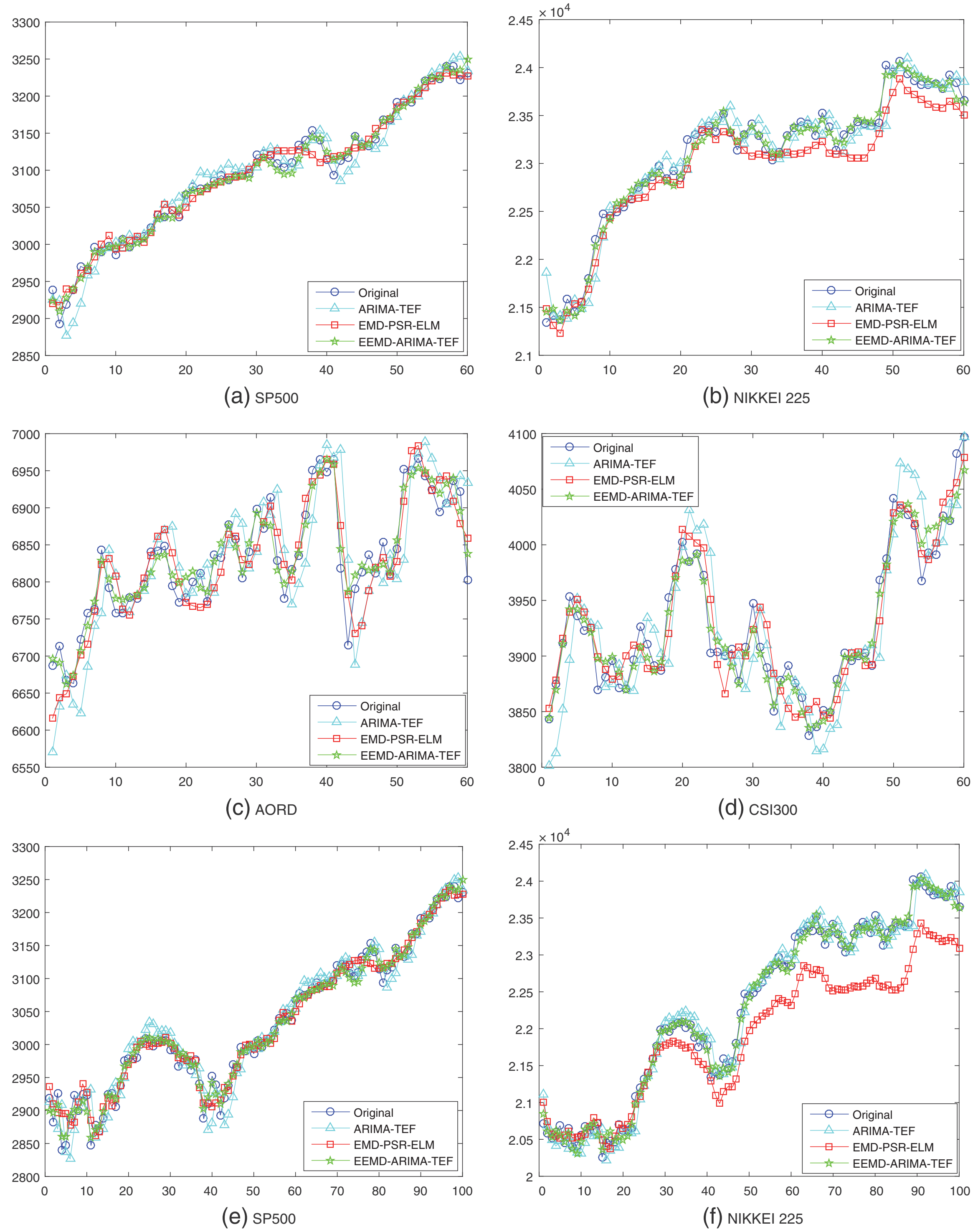
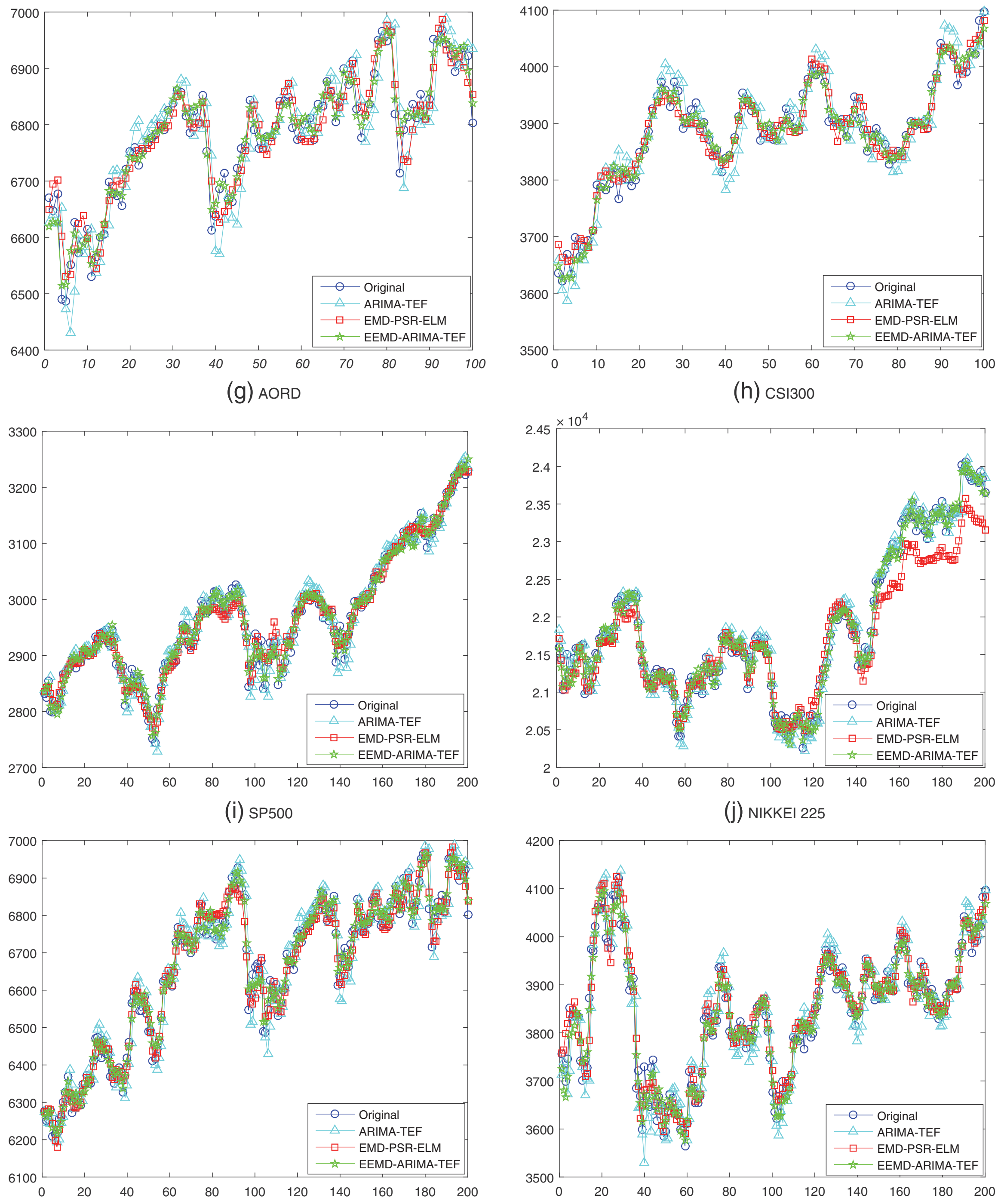


FIGURE 6 Comparison of the predicted results of closing price. Note: (a)–(d) represent predicted results of the last 60 days, (e)–(h) represent predicted results of the last 100 days, (i)–(l) represent predicted results of the last 200 days

ARIMA-TEF model and the EMD-PSR-ELM model in terms of the four criteria. The EEMD-ARIMA-TEF model outperforms the other two models for the four stock indexes for different testing data sets, which implies our EEMD-ARIMA-TEF model is quite robust to producing more accurate forecasting results.

The explanations of the proposed EEMD-ARIMA-TEF model's superiority to the ARIMA-TEF and EMD-PSR-ELM models could be summarized in two aspects: on the one hand, the EEMD is conducive to financial time series forecasting, just as previous studies and above analysis. Table 2 shows that the ARIMA-TEF model, without data preprocessing, gives the worst results among the three models. On the other hand,



(g) AORD (h) CSI300 (i) SP500 (j) NIKKEI 225 (k) AORD (l) CSI300

FIGURE 6 (Continued)

adopting the hybrid model with linear and nonlinear modelling can exploit the unique advantages of the individual models and improve the comprehensive analytical ability to complex time series. It can be seen from Equations (9)–(10) that the TEF model based on TD can utilize the whole previous data dynamically, which implies it can make full use of the concealed historical information. The ARIMA-TEF with EEMD yields better results than the EMD-PSR-ELM model, which means our proposed EEMD-ARIMA-TEF model outperforms the naïve random-walk, ARIMA, BPNN, ELM, EMD-ELM, and PSR-ELM models indirectly. Besides smaller deviation, our EEMD-ARIMA-TEF model owns a higher DS with almost at 80%. Thus our proposed hybrid model combining EEMD, ARIMA, and TEF shed a new light for modelling financial time series, and investors can develop an effective decision support system based on this model to improve investment efficiency in consideration of this model's excellent performance.

TABLE 3 Returns of the investment strategy

Stock index	Days	ARIMA-TEF	EEMD-ARIMA-TEF	EMD-PSR-ELM	Index itself
SP500	30	25.1602	181.4995	158.9397	110.3201
	60	149.1802	434.6094	400.7793	278.7700
	100	249.5103	874.7388	732.2888	292.6899
	200	437.1904	1,646.5295	1,485.5087	397.8401
NIKKEI 225	30	−617.5488	1,679.7500	1,280.1895	239.8594
	60	940.0605	4,304.1191	2,869.5977	1878.0097
	100	1,152.2285	6,768.8164	1951.1914	2,569.4589
	200	−1,010.7343	12,342.8145	7,482.3184	1930.3398
AORD	30	−505.5000	577.6992	518.5000	−96.5000
	60	−398.1000	1,115.4990	873.7993	165.5000
	100	−178.6997	1789.7988	1,435.2998	139.0000
	200	−236.4995	3,504.0981	2,839.5981	518.7998
CSI300	30	182.8692	363.5968	342.6067	149.5429
	60	133.9387	747.1130	659.9909	258.9030
	100	333.4838	1,355.3212	1,160.8464	420.8937
	200	414.8183	3,028.2341	2,785.8673	366.6275

Abbreviations: ARIMA, autoregressive integrated moving average; EEMD, ensemble empirical mode decomposition; TEF, Taylor expansion forecasting.

4.5 | Investment strategy

In this section, we set up an artificial investment strategy to show that our EEMD-ARIMA-TEF model is useful in reality. To simplify the application, we consider a long-short strategy that longs and holds the stock index when it is forecasted to go up and no buys or shorts otherwise. Because these stock indexes come from different countries, the transaction systems and costs are different. Thus, we do not consider the costs. The returns of n days are:

$$R(t)=\sum_{t=0}^{n-1}(y_{t+1}-y_t)g(\hat{y}_{t+1}-y_t),\quad \text{where } g(x)=\begin{cases}1, & x>0\\ 0, & x\le 0\end{cases} \tag{21}$$

where $R(t)$ are the returns, $y\,\cdot$ and $\hat{y}\,\cdot$ are the actual closing price and predicted price, respectively. $g(x)$ is a piecewise function, and its value of 1 means that the predicted price is rising. Its value of 0 indicates that the predicted index is falling. $y_{t+1}-y_t$ stands for the return from day t to day $t+1$. A bigger value of $R(t)$ shows a better forecast.

Next, we compute the returns of the four stock indexes based on the long-short strategy under the four different forecast horizons, and the results are shown in Table 3. We also give the returns of the stock indexes themselves. Table 3 displays that the returns from high to low are the EEMD-ARIMA-TEF, EMD-PSR-ELM, index itself, and ARIMA-TEF in general. The returns relying on the forecast by the proposed EEMD-ARIMA-TEF model are highest, which indicates our EEMD-ARIMA-TEF model is helpful to guide investment. Therefore, investors can develop an effective decision support system based on this model to acquire more profit.

5 | CONCLUSIONS

In this paper, we propose a novel hybrid model that integrates EEMD, ARIMA, and TEF to forecast financial time series. We have compared the effectiveness of the proposed model with the two benchmark models (ARIMA-TEF and EMD-PSR-ELM) developed in Zhang et al. (2016) and Yang and Lin (2017). We present empirical results using daily closing price data on four stock indexes (SP500, NIKKEI 225, AORD, and CSI300) to demonstrate the validity of the proposed EEMD-ARIMA-TEF model. The experimental results lead us to the following conclusions.

(1) The empirical results show that our EEMD-ARIMA-TEF model significantly outperforms the ARIMA-TEF model, which indicates that EEMD technology is conducive to financial time series forecasting.
(2) The proposed EEMD-ARIMA-TEF model benefits from the "linear and nonlinear" modelling framework, and it is superior to the excellent EMD-PSR-ELM model, which means our EEMD-ARIMA-TEF model outperforms the Naïve random-walk, ARIMA, BPNN, ELM, EMD-ELM, and PSR-ELM models indirectly.
(3) The examination of different testing data sets demonstrates that our EEMD-ARIMA-TEF model is robust.

(4) The EEMD, ARIMA, and TEF constitute the basics of our model, and the three methodologies are simple and easily conducted. Nevertheless, the proposed EEMD-ARIMA-TEF hybrid model improves prediction performance significantly, and the results based on the investment strategy indicate this model is helpful. These observations show the feasibility of EEMD-ARIMA-TEF for stock index price forecasting in stock index market trading and financial decision support systems.

Although our EEMD-ARIMA-TEF model is superior, there exist some limitations. In this study, only the daily closing price data sets are used as illustrative examples to evaluate the performance of the proposed model. As the daily data sets reflect the relatively short-term trend of the stock price index, while some investors may prefer medium- or long-term investment strategy. Thus, a hybrid model for forecasting with the weekly or even monthly data becomes a more critical issue in future research. Our proposed model can be extended to predict difficult forecasting tasks other than stock data, particularly complex time series data with diverse common characteristics. Our study uses only the daily closing prices on the stock index as input variables. Other input variables (e.g., some macroeconomic variables) and the relationships between different markets can also be considered to further enhance the performance of our prediction model.

ACKNOWLEDGEMENTS

The authors are very grateful to the anonymous reviewers for their valuable comments and suggestions to improve the overall quality of this paper. Guo acknowledges the NSFC Projects (61973084), Liu acknowledges the NSFC Projects (71871066) and the Shanghai Science and Technology Innovation Action Plan Project (19511101700), and Zhang acknowledges a partial support from NSF-DMS-2012298 and NSFC Projects (71991471).

CONFLICT OF INTEREST

The authors declare that this paper does not have any conflicts of interest.

ORCID

Qingfu Liu https://orcid.org/0000-0003-4252-2360

REFERENCES

Ahmad, M. A., & Mohd, T. I. (2017). A hybrid approach EMD-MA for short-term forecasting of daily stock market time series data. *Journal of Internet Banking and Commerce*, *22*(1), 1–10. https://doi.org/10.1063/1.4995933

Akman, E., Karaman, A. S., & Kuzey, C. (2020). Visa trial of international trade: Evidence from support vector machines and neural networks. *Journal of Management Analytics*, *7*, 1–22. https://doi.org/10.1080/23270012.2020.1731719

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, *31*(3), 307–327. https://doi.org/10.1016/0304-4076(86)90063-1

Box, G., & Jenkins, G. (1970). *Time series analysis: Forecasting and control*. San Francisco, CA: Holden-Day.

Crone, S. F., Hibon, M., & Nikolopoulos, K. (2011). Advances in forecasting with neural networks? Empirical evidence from the NN3 competition on time series prediction. *International Journal of Forecasting*, *27*(3), 635–660. https://doi.org/10.1016/j.ijforecast.2011.04.001

Guo, B., Han, J., & Xi, F. (2002). Linear tracking-differentiator and application to online estimation of the frequency of a sinusoidal signal with random noise perturbation. *International Journal of Systems Science*, *33*(5), 351–358. https://doi.org/10.1080/00207720210121771

Guo, B., & Zhao, Z. (2011). On convergence of tracking differentiator. *International Journal of Control*, *84*(4), 693–701. https://doi.org/10.1080/00207179.2011.569954

Han, J., & Wang, W. (1994). Nonlinear tracking-differentiator. *Journal of Systems Science and Mathematical Sciences*, *14*(2), 177–183 in Chinese.

Huang, N., Shen, Z., Long, S. R., Wu, M. C., Shih, H. H., Zheng, Q., … Liu, H. H. (1998). The empirical mode decomposition and the Hilbert spectrum for nonlinear and nonstationary time series analysis. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *454*(1971), 903–995. https://doi.org/10.1098/rspa.1998.0193

Huang, S. C., & Wu, T. K. (2008). Combining wavelet-based feature extractions with relevance vector machines for stock index forecasting. *Expert Systems*, *25*(2), 133–149. https://doi.org/10.1111/j.1468-0394.2008.00443.x

Kim, K. J. (2003). Financial time series forecasting using support vector machines. *Neurocomputing*, *55*(1), 307–319. https://doi.org/10.1016/S0925-2312(03)00372-2

Kock, A. B., & Teräsvirta, T. (2014). Forecasting performances of three automated modelling techniques during the economic crisis 2007-2009. *International Journal of Forecasting*, *30*(3), 616–631. https://doi.org/10.1016/j.ijforecast.2013.01.003

Kožić, I., & Sever, I. (2014). Measuring business cycles: Empirical mode de-composition of economic time series. *Economics Letters*, *123*(3), 287–290. https://doi.org/10.1016/j.econlet.2014.03.009

Liu, F., Wang, X., & Zhang, M. (2006). An application of nonlinear trace differential to simulating and forecasting in stock price. *Systems Engineering*, *24*(5), 81–87 in Chinese.

Lu, Y. (2019). Artificial intelligence: A survey on evolution, models, applications and future trends. *Journal of Management Analytics*, *6*, 1–29. https://doi.org/10.1080/23270012.2019.1570365

Pai, P., & Lin, C. (2005). A hybrid ARIMA and support vector machines model in stock price forecasting. *Omega*, *33*(6), 497–505. https://doi.org/10.1016/j.omega.2004.07.024

Plakandaras, V., Gupta, R., Gogas, P., & Papadimitriou, T. (2015). Forecasting the U. *S. Real House Price Index. Economic Modelling*, *45*, 259–267. https://doi.org/10.2139/ss-rn.2431627

Shi, S., Xu, L., & Liu, B. (1996). Applications of artificial neural networks to the nonlinear combination of forecasts. *Expert Systems*, *13*(3), 195–201. https://doi.org/10.1111/j.1468-0394.1996.tb00119.x

Shi, S., Xu, L., & Liu, B. (1999). Improving the accuracy of nonlinear combined forecasting using neural networks. *Expert Systems with Applications*, *16*, 49–54. https://doi.org/10.1016/s0957-4174(98)00030-x

Tiwari, A. K., Dar, A. B., Bhanja, N., & Gupta, R. (2016). A historical analysis of the US stock price index using empirical mode decomposition over 1791-2015. *Economics*, *10*(9), 1–15. https://doi.org/10.5018/economics-ejournal.ja.2016-9

Vafeiadis, T., Dimitriou, N., Ioannidis, D., Wotherspoon, T., Tinker, G., & Tzovaras, D. (2018). A framework for inspection of dies attachment on PCB utilizing machine learning techniques. *Journal of Management Analytics*, *5*(2), 81–94. https://doi.org/10.1080/23270012.2018.1434425

Wang, D., Wang, Y., Song, X., & Liu, Y. (2018). Coal overcapacity in China: Multiscale analysis and prediction. *Energy Economics*, *70*, 244–257. https://doi.org/10.1016/j.eneco.2018.01.004

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

Wang, S., Hu, A., Wu, Z., Liu, Y., & Bai, X. (2014). Multiscale combined model based on run-length-judgment method and its application in oil price forecasting. *Mathematical Problems in Engineering*, *2014*, 1–9. https://doi.org/10.1155/2014/513201

Wegener, C., Spreckelsen, C., Basse, T., & Mettenheim, H. (2016). Forecasting government bond yields with neural networks considering cointegration. *Journal of Forecasting*, *35*(1), 86–92. https://doi.org/10.1002/for.2385

Wu, Z., & Huang, N. (2009). Ensemble empirical mode decomposition: A noise assisted data analysis method. *Advances in Adaptive Data Analysis*, *1*(1), 1–41. https://doi.org/10.1142/S1793536909000047

Xie, C., Mao, Z., & Wang, G. J. (2015). Forecasting RMB exchange rate based on a nonlinear combination model of ARFIMA, SVM, and BPNN. *Mathematical Problems in Engineering*, *2015*(4), 1–10. https://doi.org/10.1155/2015/635345

Yang, H., & Lin, H. (2016). An integrated model combined ARIMA, EMD with SVR for stock indices forecasting. *International Journal on Artificial Intelligence Tools*, *25*(2), 1–22. https://doi.org/10.1142/S0218213016500056

Yang, H., & Lin, H. (2017). Applying the hybrid model of EMD, PSR, and ELM to exchange rates forecasting. *Computational Economics*, *49*(1), 99–116. https://doi.org/10.1007/s10614-015-9549-9

Yuan, R., Li, Z., Guan, X., & Xu, L. (2010). An SVM-based machine learning method for accurate internet traffic classification. *Information Systems Frontiers*, *12*, 149–156. https://doi.org/10.1007/s10796-008-9131-2

Zeng, Q., Qu, C., Ng, A., & Zhao, X. (2016). A new approach for Baltic dry index forecasting based on empirical mode decomposition and neural networks. *Maritime Economics & Logistics*, *18*(2), 192–210. https://doi.org/10.1057/mel.2015.2

Zhang, G. (2003). Time series forecasting using a hybrid ARIMA and neural network model. *Neurocomputing*, *50*, 159–175. https://doi.org/10.1016/S0925-2312(01)00702-0

Zhang, G., Zhang, X., & Feng, H. (2016). Forecasting financial time series using a methodology based on autoregressive integrated moving average and Taylor expansion. *Expert Systems*, *33*(5), 501–516. https://doi.org/10.1111/exsy.12164

Zhang, J., Wei, Y., Tan, Z., Wang, K., & Tian, W. (2017). A hybrid method for short-term wind speed forecasting. *Sustainability*, *9*(4), 1–10. https://doi.org/10.3390/su9040596

Zhang, N., Lin, A., & Shang, P. (2017). Multidimensional k-nearest neighbor model based on EEMD for financial time series forecasting. *Physica A*, *477*, 161–173. https://doi.org/10.1016/j.physa.2017.02.072

Zhang, X., Lai, K. K., & Wang, S. Y. (2008). A new approach for crude oil price analysis based on empirical mode decomposition. *Energy Economics*, *30*(3), 905–918. https://doi.org/10.1016/j.eneco.2007.02.012

Zhu, B., Shi, X., Chevallier, J., Wang, P., & Wei, Y. (2016). An adaptive multiscale ensemble learning paradigm for nonstationary and nonlinear energy price time series forecasting. *Journal of Forecasting*, *35*(7), 633–651. https://doi.org/10.1002/for.2395

AUTHOR BIOGRAPHIES

Zhidan Luo is a PhD candidate in the School of Statistics at University of International Business and Economics. His research interests are in the areas of financial time series forecasting and data driven control and learning system.

Wei Guo received the PhD degree from the Academy of Mathematics and Systems Science, the Chinese Academy of Science, Beijing, China, in 2004. He worked in the School of Computational and Applied Mathematics, University of the Witwatersrand, South Africa in 2007 as a Postdoctoral Fellow. He is currently a Professor with the University of International Business and Economics, Beijing. His research interests include theory of control and application and data driven control and learning system.

Qingfu Liu received the PhD degree in Management Science and Engineering from Southeast University, Nanjing, China, in 2005. He worked in the School of Economics, Fudan University, Shanghai in 2007 as a Postdoctoral Fellow. He is currently a professor in the Institute for Financial Studies at Fudan University. His research interests cover big data finance, risk management and financial technology.

Zhengjun Zhang received the PhD degree in Statistics from the University of North Carolina, Chapel Hill, USA, in 2002. He is currently a professor in the Department of statistics at University of Wisconsin-Madison. His research interests include financial time series analysis, extreme value theory, market systematic risk assessment and financial risk modelling and evaluation. He has published dozens of papers in these areas.

How to cite this article: Luo Z, Guo W, Liu Q, Zhang Z. A hybrid model for financial time-series forecasting based on mixed methodologies. *Expert Systems*. 2020;e12633. https://doi.org/10.1111/exsy.12633