



*Research article*

## **A hybrid model combining variational mode decomposition and an attention-GRU network for stock price index forecasting**

**Hongli Niu\* and Kunliang Xu**

School of Economics and Management, University of Science and Technology Beijing, Beijing 100083, China

\* **Correspondence:** Email: niuhongli@ustb.edu.cn.

**Abstract:** In this paper we introduce a new hybrid model based on variational mode decomposition (VMD) and Gated Recurrent Units (GRU) network improved by attention mechanism to enhance the accuracy of stock price indices forecasting. In the process of establishing the model, VMD is made a use to decompose the primary series into some almost orthogonal subsequences. The attention mechanism is introduced into GRU to assign different weights to the input elements in advance so that better predictive results can be achieved for each component. In empirical experiment, London FTSE Index (FTSE) and Nasdaq Index (IXIC) are adopted to examine the performance of VMD-AttGRU model. Empirical results report that the developed hybrid model outperforms the single models and indeed raises the accuracy of stock price indices forecasting. In addition, the introduction of attention mechanism can increase the level predictive accuracy but decrease the correctness of direction forecasting.

**Keywords:** variational mode decomposition; Gated Recurrent Units; attention mechanism; forecasting; stock price

---

### **1. Introduction**

As stock markets gradually enter the public vision, the precise prediction of stock price indices has become one of the most promising research projects in forecasting of time series. The commonly used forecasting methods are simply divided into two classes: econometric methods and artificial intelligence (AI) based models. The latter, represented by artificial neural networks (ANNs), have been proved to outperform the econometric methods in dealing with non-stationary and non-linear time

series [1–4]. As an improvement of traditional ANN, recurrent neural networks (RNN) [5] establish connections between the hidden layer units, through which the dependency of data at different time points can be further explained. The before-after associated structure ensures that RNN is especially suitable for predicting time series data [6]. By introducing three gate mechanism into the hidden units of traditional RNN, long short-term memory (LSTM) network overcomes the short comings existing in RNN, such as gradient disappearing and exploding in long time span [7]. Recently, LSTM has been widely utilized to predict time series and obtained outstanding results [8,9]. Gated recurrent units (GRU) network integrates the three gates of LSTM into reset gate and update gate, which effectively improves the computing efficiency of LSTM [10], and GRU achieved better results than LSTM in different time series forecasting tasks [11,12]. In this paper, the attention mechanism is introduced to assign weights to different input elements of GRU and obtain a more precise forecasting result.

To further improve the forecasting accuracy of stock price indices, hybrid models containing two or more individual models have been developed gradually, in which the unique advantages of different individual models can be exploited. Following “Divide-and-Conquer” principle, “Decomposition-and-Ensemble” is a typical framework employed in time series forecasting [13], the main idea of which is to decompose an raw complex sequence into several subseries with simple patterns so as to establish a prediction model for every subseries, and the final result is concluded by summing up the prediction results of the subseries [14]. Based on the excellent performance, hybrid forecasting models are becoming the mainstream gradually [15]. As a novel multiresolution technique originated from signal processing, variational mode decomposition (VMD) [16] is a completely non-recursive algorithm that can decompose the original series into multiple components with a specific bandwidth in the spectral domain. It has been proved that VMD performs better than the models of the same kind, such as Empirical mode decomposition (EMD) [14], in noise robustness and component decomposing accurately. In recent years, the hybrid models based on VMD have been applied successfully in several fields. For instance, by integrating the VMD with classical ANNs, Lahmiri [17] established a forecasting model VMD-PSO-BPNN for intraday stock prices prediction. The experimental results in terms of six stocks suggests that the hybrid model performs better than the single PSO-BPNN model significantly. However, there is no methodology regarding optimal selection of the number of subcomponents of VMD. In his follow-up research [18], the newly proposed model VMD-GRNN demonstrates higher accuracy than the EMD-based forecasting models in the predictions of WTI oil prices, CANUS exchange rate and NASDAQ 100 VIX when the parameter of subcomponents number ranges from 6 to 12. The similar results are proved in [19], in which the VMD is combined with a GRNN optimized by particle swarm optimization (PSO) and the hybrid model is established to predict the California electricity and Brent crude oil prices. The performances of EMD and VMD-based models are assessed and the number of subseries of VMD is set to be the same as EMD. The above researches have confirmed the applicability and superiority of VMD in practice, but it still has some room for improvement: Firstly, the optimal number of components decomposed by VMD is still difficult to be determined, but the empirical results of literature [18] have indicated that the forecasting quality of the VMD-based models will vary with the change of component number decomposed. Secondly, the above-mentioned forecasting models are all classical ANNs, which can be replaced with the promising RNNs, such as RNN, LSTM and GRU, to further enhance the forecasting ability. Thirdly, the evaluating metrics are only limited to error measures without considering the capability of correctly predicting the moving direction of the time series, which is of great significance in the short-term prediction of financial time series data.

Combining the advantages of GRU, VMD network and other variant models, there have been several literatures utilizing the hybrid forecasting models to implement various prediction tasks. For example, Zhu et al. [20] employed a hybrid model integrating VMD and BiGRU network to forecast the daily natural rubber futures price and volatility, validating the effectiveness of this model. The result indicated that the improvements in prediction performance largely depended on the time-scale matching degree between the predicted target and the mode sub-series. Li et al. [21] introduced an error correction strategy into VMD-GRU hybrid model to enhance the model performance in wind speed interval prediction, and the experiments based on eight cases from two wind fields demonstrated the proposed model is a highly qualified forecasting method. By combining GRU with VMD, Wang et al. [22] adopted a hybrid model for addressing the wind power interval prediction problem and proposed an optimization method based on constructed intervals for building high-quality training labels before applying the Adam algorithm for full training, and the effectiveness of the VMD-GRU was confirmed in comparison with other models. However, it is worth noting that the historical elements input into the forecasting network play different roles when predicting the target value in time series. In general, the impact of the input values closer to the target value is greater than that of the farther time points. Moreover, the optimal number of components needs to be preset in VMD, which is important to improve the accuracy of the final prediction result. In this work, after decomposing the original time series into an optimal number of subseries according to a certain standard the ratio of residual energy ( $r_{res}$ ) by VMD, an attention mechanism is introduced into the GRU network to enhance the forecasting quality by assigning different weights to the input elements.

The contribution of this paper to the literature is to propose a novel hybrid model for the reliable stock price indices time series prediction, namely, London FTSE Index (FTSE) and Nasdaq Index (IXIC). The evaluations indicate that compared with the counterparts, including the single models and the traditional GRU-based models, the proposed VMD-AttGRU model presents more accurate and robust results demonstrated by the level forecasting indices. The introduction of attention mechanism in the hybrid model VMD-GRU decreases the forecasting error while slightly reduce the ability of this model to correctly predict the direction.

## 2. Methodologies

### 2.1. Variational mode decomposition (VMD)

Variational mode decomposition (VMD) is a non-recursive and adaptive data decomposition technique developed recently [16]. VMD is utilized, in the VMD-AttGRU model, to decompose the original stock index  $x(t), t = 1, 2, \dots, N$  into  $n$  components,  $c_i, i = 1, 2, \dots, n$ , which stands for different local vibrations ranging from high frequency to low frequency. Each mode  $c_i$  need to compact around a center frequency  $\omega_k$  mostly. The bandwidth of a mode can be estimated by follows: At first, for each mode  $c_i$ , the Hilbert transform is employed to calculate the correlation analysis data and a unilateral frequency spectrum is obtained. Then, for each mode  $c_i$ , the spectrum of mode is transmitted to the baseband by exponential mixing with the pulses tuned to their respective centers. Afterwards, the  $H^1$  Gaussian smoothness of the demodulated series is used to calculate the bandwidth. The constraints of variational problem can be expressed in the following way:

$$\min_{(\omega_k, u_k)} \left\{ \sum_{i=1}^n \left\| \partial_t \left[ \left( \delta_t + \frac{j}{\pi t} \right) * c_i(t) \right] e^{-j\omega_i t} \right\|_2^2 \right\} \quad (1)$$

$$s. t. \quad \sum_{i=1}^n c_i = x(t)$$

where  $\{c_i\} = \{c_1, c_2, \dots, c_n\}$  and  $\{\omega_i\} = \{\omega_1, \omega_2, \dots, \omega_n\}$  respectively denote the set of the  $i_{th}$  subcomponent and its corresponding central frequency.  $\partial_t$  indicates the differential processing of  $t$ ,  $\|\cdot\|$  indicates the norm processing,  $\delta_t$  represents the Dirac function, and  $*$  denotes the convolution symbol.

To solve the optimization problem of constrained variational decomposition, an augmented Lagrangian function  $L$  is introduced:

$$L(\{c_i\}, \{\omega_i\}, \lambda) = \alpha \sum_{i=1}^n \left\| \partial_t \left[ \left( \delta_t + \frac{j}{\pi t} \right) * c_i(t) \right] e^{-j\omega_i t} \right\|_2^2 \quad (2)$$

$$+ \left\| x(t) - \sum_{i=1}^n c_i(t) \right\|_2^2 + \left\langle \lambda(t), x(t) - \sum_{i=1}^n c_i(t) \right\rangle,$$

in which  $\alpha$  denotes the penalty parameter, and  $\lambda(t)$  is the Lagrangian multiplier. In order to obtain the saddle point of the above formula, which also is the solution of the original constraint conditions, VMD adopts the Alternate direction method of multipliers (ADMM) [23].

Prior to VMD, the number of components  $n$  should be properly determined in advance. If the number is large, additional computing resources will be occupied, but if  $n$  is small, it may lead to an insufficient decomposition and inaccurate forecasting results finally. The ratio of residual energy  $r_{res}$  to original data sequence energy is used to determine the optimal number, which can be formulated as follows:

$$r_{res} = \frac{1}{N} \sum_{t=1}^N \left| \frac{x(t) - \sum_{i=1}^n c_i(t)}{x(t)} \right| \quad (3)$$

where  $r_{res}$  is the residual after decomposition, which can be used as the optimization index of VMD process. In empirical, when  $r_{res}$  is smaller than 1% or there is no obvious trend of downwards, the component number can be defined [24].

## 2.2. Long short-term memory network and gated recurrent unit network

The long short-term memory (LSTM) network [8] creatively introduces the “gate” mechanism to improve the conventional recurrent neural network (RNN): it replaces the hidden layer nodes of the RNN with special memory cells. Each memory cell contains three gates: input gate  $i_t$ , forget gate  $f_t$ , and output gate  $o_t$  that implement the filtering and processing of historical states and information, and the problems of gradient disappearance and explosion can be effectively resolved. The LSTM has been successfully applied in time series prediction [8,9]. The gated recurrent unit (GRU) network [10] integrates the three gates of the LSTM into two gates: reset gate  $r_t$  and update gate  $z_t$  and achieves better performance in time series forecasting tasks [25]. The reset gate measures how much the historical information will be kept at this moment and how much the latest information will be added,

which helps to grasp the dependency of short-term existing in the series data, while the update gate determines the degree of “forgetting” historical information, and the information with arbitrary-lengths of the input  $x_t$  can be memorized in this gate effectively. The basic steps of GRU can be shown in the following:

At first, the reset gate  $r_t$  and update gate  $z_t$  at the current state (time  $t$ ) are established by the latest input  $x_t$  and the hidden state produced by the previous cell  $h_{t-1}$ , and the outputs of the two gates are respectively given as:

$$r_t = \sigma(x_t U^r + h_{t-1} W^r + b^r) \quad (4)$$

$$z_t = \sigma(x_t U^z + h_{t-1} W^z + b^z) \quad (5)$$

Secondly, the current candidate hidden state  $\tilde{h}_t$  can be formulated:

$$\tilde{h}_t = \tanh(x_t U^h + (h_{t-1} * r_t) W^h + b^h) \quad (6)$$

Finally, the outcome of current hidden state  $h_t$  can be computed by implementing the linear combination of the current candidate hidden state  $\tilde{h}_t$  and the previous hidden state  $h_{t-1}$ , where the sum of weighting coefficient is equal to 1.

$$h_t = (1 - z_t) * \tilde{h}_t + z_t * h_{t-1} \quad (7)$$

where  $U^r, U^z, U^h$  and  $W^r, W^z, W^h$  represent the appropriate weight coefficient matrices,  $b^r, b^z$  and  $b^h$  denote the corresponding bias vectors,  $\sigma(\cdot)$  and  $\tanh(\cdot)$  are the Sigmoid function and Hyperbolic tangent function respectively, and  $*$  indicates the dot multiplication between matrices.

### 2.3. Attention mechanism

Attention mechanism is originated from a fact that human brain focuses on only specific parts of their visual view when recognizing something [26]. For predicting time series, there is a fact that not all elements in the input series contribute equally to the value of context vector at each time step  $t$ , which is often ignored by the conventional forecasting networks. Therefore, the principle of attention mechanism built in neural network is to select crucial elements and give more weight to them, rather than taking all elements into account equally. That is, the attention mechanism is a deep learning algorithm for identifying the most relevant inputs. After ignoring the irrelevant information and amplifying the needed information, the processing efficiency of input information is greatly improved. Recently, the attention mechanism has been applied in computational neuroscience [27], text representations [28] and image description [29] successfully. Figure 1 depicts the calculation of attention value in three steps, through which different weights  $w_i$  are assigned to the elements of input series to highlight the important subset of its inputs by training the model at different time. Every element of the input data set is assumed to contain an address (Key) and a value (Value). The given goal is denoted as  $G$  and the attention weight is the result to be calculated. In the figure,  $F(G, \text{Key})$  is adopted to calculate the relevancy between the given target  $G$  and address  $K$ .  $R_i$  and  $w_i$  ( $i = 1, 2, \dots, m$ ) represent respectively the relevance and weight of attention for the  $i_{th}$  element of input sequence at time  $t$ . The realization of attention mechanism can be formulated as follows:

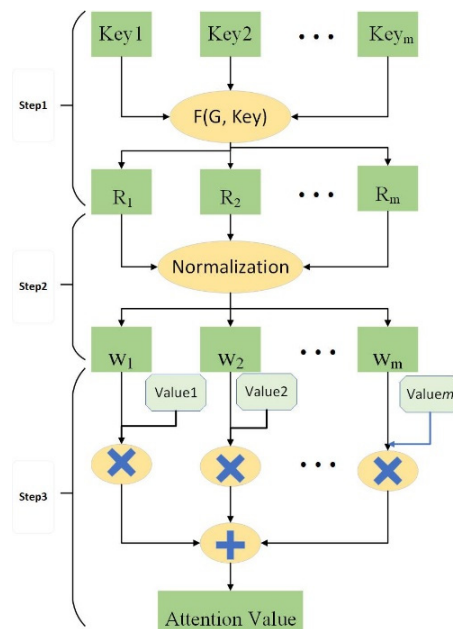
$$e_t = \text{Attend}(x_t, s_{t-1}, w_{t-1}) \quad (8)$$

$$w_{tj} = \frac{\exp(e_{tj})}{\sum_{j=1}^N \exp(e_{tj})} \quad (9)$$

$$\hat{x}_{tj} = w_{tj}x_{tj} \quad (10)$$

where  $e_t$  denotes the attention score that is defined by input data  $x_t$ , previous state  $s_{t-1}$  and weight  $w_{t-1}$  of previous attention.

The specific implementation process of attention mechanism utilized in this work is referred to [30]. That is, in the first step, the relevancy between every previous input elements and output elements are computed. Then, applying the softmax formula to convert the relevancies into the probability form. Lastly in the third step, multiply the obtained probabilities by the implicit expression of the corresponding input feature, to make it stand for the feature contribution to the forecasted load and sum up all the input contribution features to be the input section to forecast the next load value.



**Figure 1.** Three steps of attention value calculation.

#### 2.4. VMD-AttGRU network

In view of the advantages of VMD, attention mechanism and GRU network, we construct a hybrid model named VMD-AttGRU by combining the three techniques. In this model, the VMD is utilized to decompose the original time series into several components. The Attention-GRU (simplified as AttGRU) is used to establish forecasting model for each component and obtain the predicted output separately, in which the GRU layer takes the output of the attention layer as the input so that the capability of conventional GRU network is improved. The final forecasting result is calculated by summarizing the separate predicted outputs obtained by AttGRU. The flow chart in Figure 2 depicts its implementation process, in which the VMD-AttGRU operation is carried out as follows.

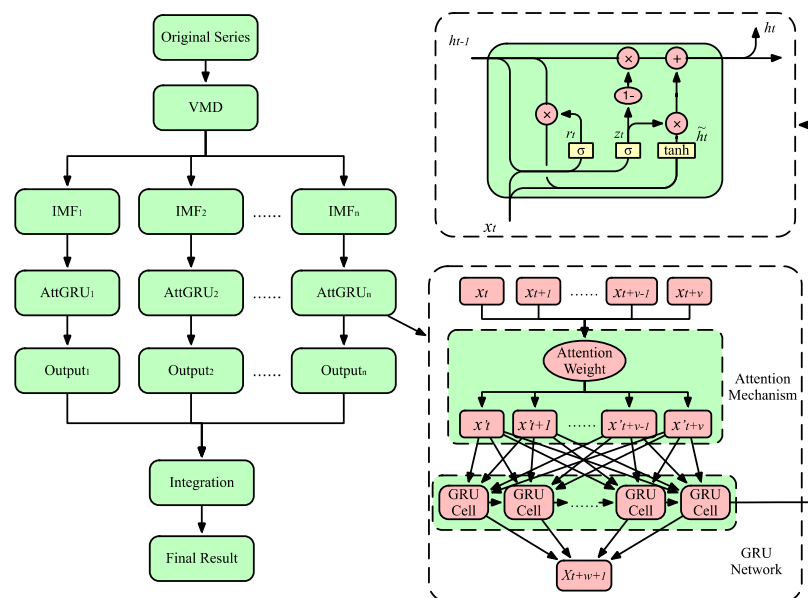
Step 1: The VMD is utilized to decompose the stock price index series  $x(t)$ ,  $t = 1, 2, \dots, N$ , into  $n$  mutually independent subseries, denoted by  $IMF_1, IMF_2, \dots, IMF_n$ , in which the  $n$  is determined by a specific standard. The initial series is reconstructed in terms of the IMFs as:

$$x(t) = \sum_{k=1}^n IMF_k(t)$$

Step 2: Each component IMF is split into training and test datasets at a fixed ratio, and the input and output sets are split according to the step size. The AttGRU network is utilized to train and establish the forecasting model based on the training dataset. The forecasting output of each IMF is obtained.

Step 3: The final predicted result of the original stock price index series is calculated by summarizing the separate predicted outputs.

Step 4: Multiple performance measures, i.e., MAE, RMSE, MAPE, TIC, and  $D_{stat}$ , are adopted to evaluate the prediction capacity of VMD-AttGRU from different perspectives.



**Figure 2.** The structure of VMD-AttGRU.

### 3. Data selection and processing

In this work, the daily closing price of London FTSE Index (FTSE) and Nasdaq Index (IXIC) are used to examine the validity of the proposed VMD-AttGRU model. The selected two stock price indices are both representative in the global stock markets and regarded as important benchmarks of social and economic development. They are collected from the global important stock price indices of Wind database, which stored in the form of [date, price] time series. The FTSE cover the time period from 2007/03/09 to 2020/06/05, which accounts 3348 data points, and the IXIC cover the time period from 2007/02/20 to 2020/06/05, which also account 3348 data points. To conduct experiments, the first 80% of each sample is used to train the model, and the remaining 20% is used as test sets. Figure 3 displays the curves of price samples of FTSE and IXIC. Table 1 exhibits the details illustration of the

selected two stock price indices. Table 2 shows the descriptive statistic information of the samples in terms of mean, standard deviation, skewness, kurtosis, Jarque-Bera (JB) test for normality and Augmented Dickey Fuller (ADF) test for stationarity. It is shown that with the standard deviation value of 2114.71 for IXIC and 894.81 for FTSE, the IXIC has more volatility than the FTSE. The FTSE is negatively skewed with skewness value of  $-0.56$  while the IXIC is positively skewed with skewness of  $0.67$ . Both of them have kurtosis less than 3, implying no leptokurtosis. The results of JB test indicate that both FTSE and IXIC price index series are distinctly non-Gaussian distributed at the 5% confidence level. The results of ADF test suggests the significantly non-stationary of both prices.

To reduce the impact of noise and facilitate optimize the solving process, each component  $c(t), t = 1, 2, \dots, N$  obtained by VMD will be normalized to the range of  $[0, 1]$  by the following maximum and minimum standardized formula:

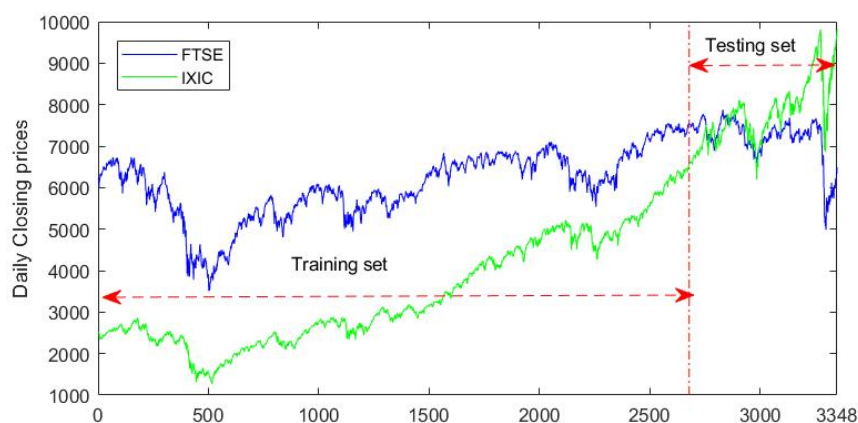
$$c(t)' = \frac{c(t) - \min c(t)}{\max c(t) - \min c(t)} \quad (11)$$

Then the normalized data is input into the AttGRU network for training and prediction. In order to obtain the real predictive value and compare it with the actual value intuitively the normalized output  $c'(t)$  can be reverted to  $x(t)$  after prediction as follows:

$$c(t) = c'(t)(\max c(t) - \min c(t)) + \min c(t) \quad (12)$$

**Table 1.** Datasets of the selected stock price indices for forecasting.

	FTSE	IXIC
Time period	2007/03/09 ~ 2020/06/05	2007/02/20 ~ 2020/06/05
Total number	3348	3348
Train sets	2007/03/09 ~ 2017/10/13	2007/02/20 ~ 2017/10/05
Train number	2678	2678
Test sets	2017/10/16 ~ 2020/06/05	2017/10/06 ~ 2020/06/05
Test number	670	670



**Figure 3.** Daily closing prices of FTSE and IXIC stock indices.



**Table 2.** Descriptive statistics of the FTSE and IXIC.

Index	Mean	Std.	Skewness	Kurtosis	JB test	ADF test
FTSE	6273.01	894.81	-0.56	2.90	177.20* (0.00)	-0.25 (0.56)
IXIC	4320.50	2114.71	0.67	2.28	323.25 *(0.00)	2.13 (0.99)

#### 4. Performance evaluation metric

We would like to better validate the robustness of the prediction network of VMD-AttGRU, this work adopts five commonly-used criteria to examine the superiority of the model from the various perspectives. They are including the mean absolute error (MAE), root mean square error (RMSE), mean absolute percentage error (MAPE), Theil Inequality Coefficient (TIC) and directional statistic  $D_{stat}$ , in which the first four indices are employed to measure the level forecasting accuracy and the  $D_{stat}$  is employed to measure the correctness of predicted direction for a time series in terms of percentage. They are respectively defined as follows:

$$MAE = \frac{1}{N} \sum_{t=1}^N |x_t - \hat{x}_t| \quad (13)$$

$$RMSE = \sqrt{\frac{1}{N} \sum_{t=1}^N (x_t - \hat{x}_t)^2} \quad (14)$$

$$MAPE = \frac{1}{N} \sum_{t=1}^N \left| \frac{x_t - \hat{x}_t}{x_t} \right| \quad (15)$$

$$TIC = \frac{\sqrt{\frac{1}{N} \sum_{t=1}^N (x_t - \hat{x}_t)^2}}{\sqrt{\frac{1}{N} \sum_{t=1}^N x_t^2} + \sqrt{\frac{1}{N} \sum_{t=1}^N \hat{x}_t^2}} \quad (16)$$

$$D_{stat} = 100\% * \frac{1}{N} \sum_{t=1}^N d_t \quad (17)$$

$$d_t = \begin{cases} 1, & \text{if } (\hat{x}_t - \hat{x}_{t-1})(x_t - x_{t-1}) \geq 0, t > 1 \\ 0, & \text{otherwise} \end{cases}$$

where  $x_t$  expresses the actual value,  $\hat{x}_t$  signifies the forecasting value,  $N$  is the length of sample of forecasting results, the same applies hereinafter. The MAE is used to measure the average absolute error between the actual series and the predicted series. The RMSE, which is more sensitive to outliers, is used to measure the deviation between the actual and the predicted series. The MAPE is designed to compute the average relative errors between the actual series and the predicted series in terms of percentage, while the directional statistic  $D_{stat}$  is adopted to evaluate the capability of correctly predicting the moving direction of the time series. In general, the smaller value of the MAE, RMSE, MAPE and TIC indicates the less difference between the forecasting and the actual values, that is, the

more accuracy of the prediction of the model. The higher  $D_{stat}$  value corresponds to the better performance of the model.

## 5. Empirical results

In this section, the predictive performance of VMD-AttGRU model for stock price indices forecasting is analyzed. To comprehensively demonstrate the advantages of the proposed hybrid model and the effectiveness of the attention mechanism in stock price index prediction, single models (LSTM, GRU, AttGRU) and hybrid model VMD-GRU are considered for comparison. According to the “decomposition and ensemble” strategy, at first the prices are decomposed by VMD technique, in which the number of subseries IMFs should be determined first. Table 3 displays the ratio of residual energy  $r_{res}$  in VMD approach under different  $n$  for the stock price indices. All  $r_{res}$  are below 1%. In FTSE, the downward tendency of  $r_{res}$  tends to be stable when  $n$  is larger than 15, while the descending tendency of  $r_{res}$  tend to be stable when  $n$  is larger than 16 in IXIC. Therefore, the suitable number of components in FTSE is set 15, and that in IXIC is set 16.

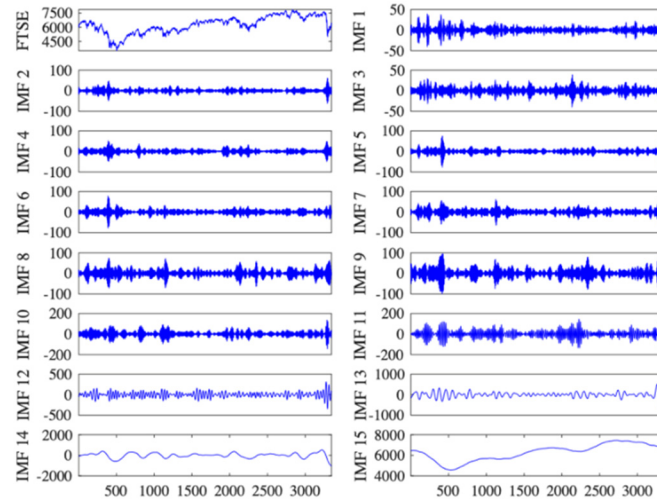
Taking FTSE as an example, Figure 4 displays the subseries obtained by VMD. They are listed ranging from high to low frequency, depicting different local oscillations embodied in the data series. It can be seen intuitively that the decomposed subseries is more regular than the original series, which helps to reduce the complexity of datasets to be forecasted. Among them, the high frequency components with relatively small values reflecting the detailed short-term volatilities information of the original price series, and the low frequency components composed of large values represent the whole changes of tendency of the daily closing prices.

**Table 3.** The ratio of residual energy under different  $n$ .

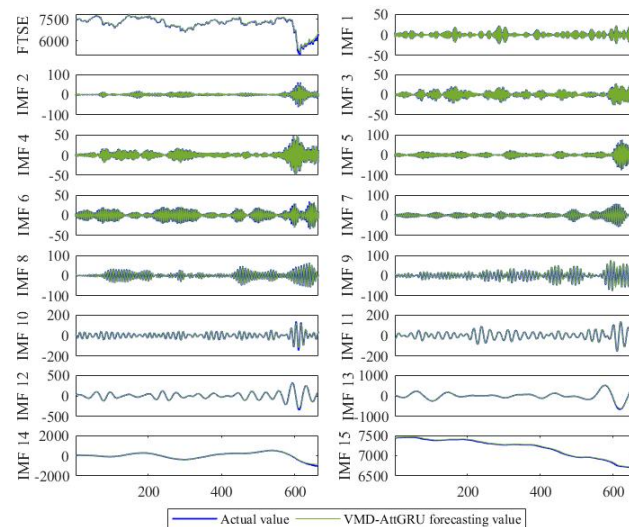
$n$	FTSE	IXIC
5	0.54%	0.68%
6	0.45%	0.54%
7	0.40%	0.42%
8	0.36%	0.34%
9	0.31%	0.30%
10	0.27%	0.28%
11	0.23%	0.25%
12	0.20%	0.23%
13	0.17%	0.22%
14	0.15%	0.21%
15	<b>0.13%</b>	0.17%
16	0.12%	<b>0.14%</b>
17	0.11%	0.14%
18	0.10%	0.13%

Later, the corresponding AttGRU prediction model is constructed for each composed IMF subseries. In parameters setting, A historical lag of order 5 is taken to predict the data of the next period, considering there are 5 trading days per week that can be regarded as a cycle simply. In other word, the number of input data points is set to 5 and that of outputs is set to 1. After repeated

experiments, a  $5 \times 50 \times 1$  neural network is obtained by setting the number of hidden nodes to 50. For convenience, set the number of epochs to 300 and the batch size to 64. It should be noted that all of the processes are implemented in Python 3.x running on a Quad-Core Intel Core i5 processor operating at 1.40 GHz with an 8 GB installed RAM.



**Figure 4.** The subseries of FTSE obtained by VMD.

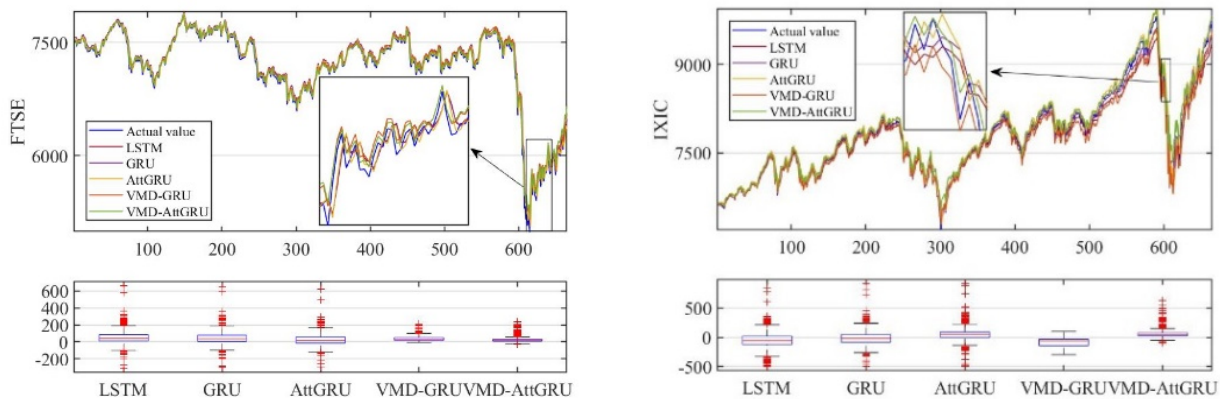


**Figure 5.** The AttGRU forecasting results for each subseries for FTSE data in the test set.

Figure 5 shows the comparison of the actual value and forecasted value by AttGRU for each subseries in the FTSE test set. It shows that the predicted curve is very close to the real curve of each subseries, demonstrating that the AttGRU network can make an accurate prediction of components with different frequency information.

Figure 6 shows the results for VMD-AttGRU for the two stock price indices test sets along with the other considered models: LSTM, GRU, AttGRU, and VMD-GRU. Overall, for both stocks, the

curves are close together, showing that the predicted curve of each model is near the real price curve. The curve for the VMD-AttGRU model is generally the closest to the actual curve, indicating the best prediction performance in this comparison. This can be further observed in the inset plots, where a certain volatile part of the datasets is magnified. So, we can conclude that the VMD-AttGRU model has the highest accuracy for stock price prediction.



**Figure 6.** Forecasting results and error of different models for the stock price indices.

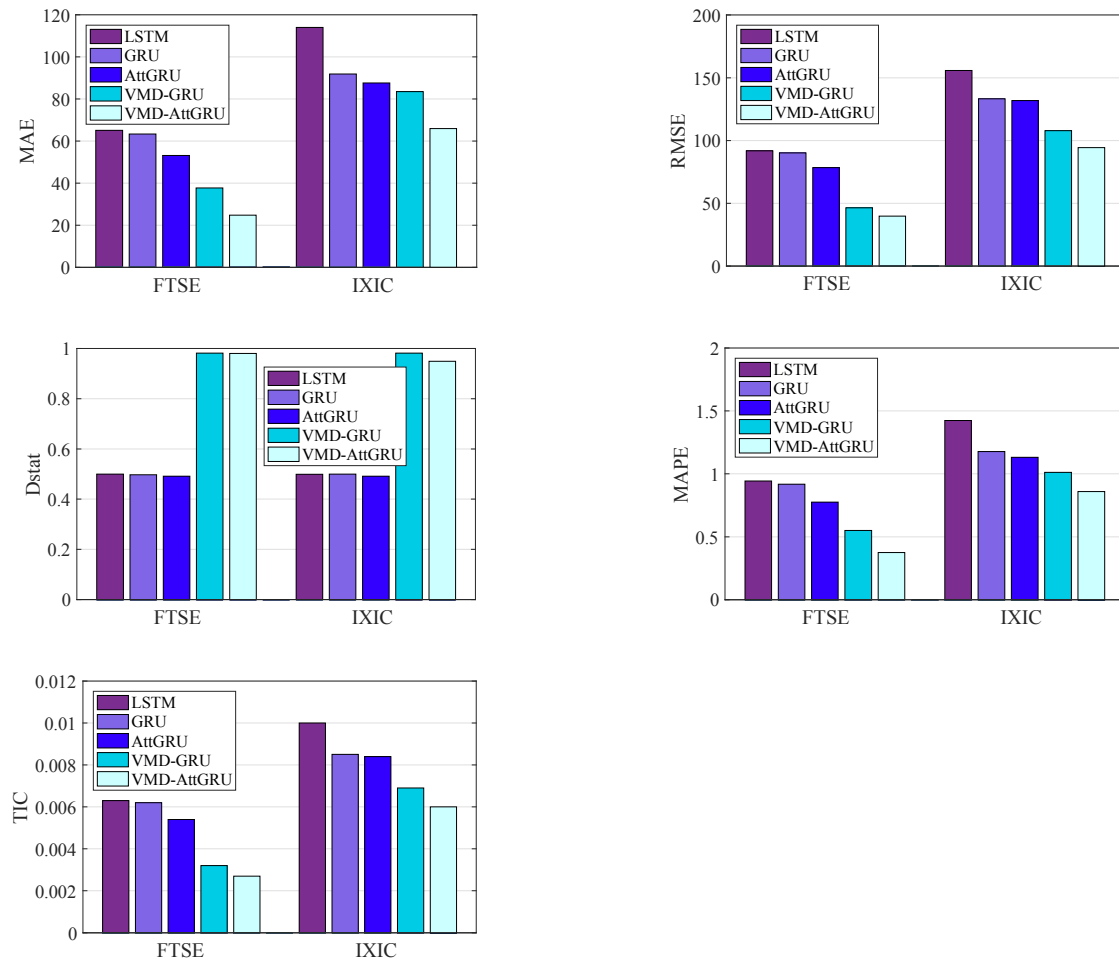
**Table 4.** Predictive performance of different models for FTSE data.

Models	MAE	RMSE	MAPE (%)	TIC	$D_{stat}$	Time (s)
LSTM	65.114	91.666	0.943	0.0063	50.00%	42.136
GRU	63.334	90.061	0.918	0.0062	49.70%	45.844
AttGRU	53.185	78.302	0.776	0.0054	49.10%	48.012
VMD-GRU	37.725	46.339	0.551	0.0032	98.19%	687.125
VMD-AttGRU	24.802	39.683	0.375	0.0027	98.04%	744.332

**Table 5.** Predictive performance of different models for IXIC data.

Models	MAE	RMSE	MAPE (%)	TIC	$D_{stat}$	Time (s)
LSTM	113.998	155.707	1.423	0.0100	49.85%	42.887
GRU	91.874	133.213	1.176	0.0085	50.00%	45.878
AttGRU	87.607	131.814	1.132	0.0084	49.10%	49.032
VMD-GRU	83.503	107.762	1.012	0.0069	98.19%	731.371
VMD-AttGRU	65.925	94.245	0.858	0.0060	94.88%	798.146

In order to further analyze the performance of various models, the predictive errors are also presented in Figure 4. It can be seen that the upper and lower bounds are not much different for single models. The prediction errors of single models are evidently larger than those of the hybrid model. The median of the VMD-AttGRU model is closest to 0, and the absolute values of the upper and lower quartiles are the smallest in the comparison group. The results further show that the relative error of the target model is relatively smaller and more concentrated, illustrating the better performance of the proposed model in stock price series data.



**Figure 7.** Forecasting performance evaluation of different models for FTSE and IXIC data.

To quantitatively measure the predictive performance of each model, the evaluation criteria MAE, RMSE, MAPE, TIC,  $D_{stat}$  and processing time are calculated in Tables 4 and 5, and the bar graphs are given in Figure 7. It can be observed that:

1) The hybrid forecasting models following the decomposition-and-ensemble strategy outperform the single models comprehensively, especially for the directional statistic  $D_{stat}$ , which is approximately at a level of 50% in single models but is improved by more than 40% after combining with VMD. For error-type performance measures including MAE, RMSE, MAPE, and TIC, the values of the VMD-based models are all smaller than single models, which also verifies the superior performance of the hybrid models in stock price index forecasting.

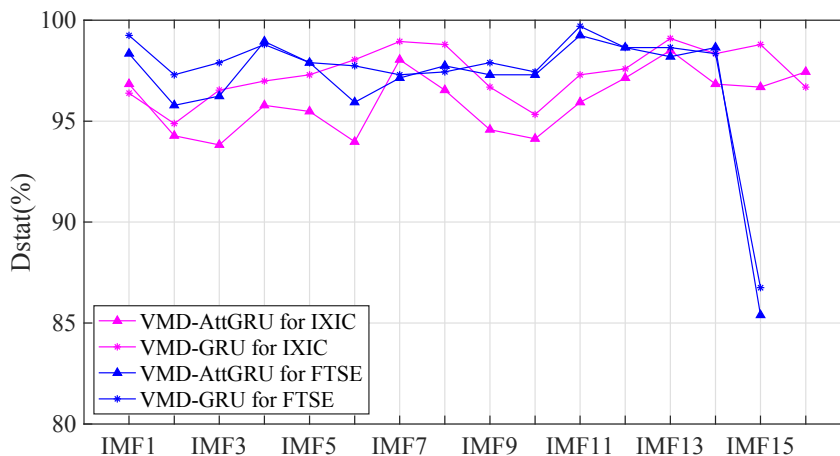
2) When introducing the attention mechanism to the GRU network, the error-type performance measures obviously decrease, indicating an improvement of forecasting accuracy. Taking MAPE for the FTSE data as an example, the MAPE of GRU is 0.918, while that of AttGRU is 0.776, reduced by 15.46%. VMD-GRU has a MAPE value of 0.551 and VMD-AttGRU has a value of 0.375, reduced by 51.57%. However, the accuracy measured by  $D_{stat}$  decreases for both FTSE and IXIC data after adding the attention mechanism. Specifically, the value for AttGRU and VMD-AttGRU is smaller than that for GRU and VMD-GRU respectively. Considering that the final predicted result is determined by the linear summation of predicted results of different IMFs and the forecasting quality of each IMF

affects the final result largely, Figure 8 further exhibits the comparison of  $D_{stat}$  value of each IMF predicted by AttGRU and GRU for the VMD-based hybrid model, in which the  $D_{stat}$  values of different IMFs predicted by GRU are generally higher than that predicted by AttGRU for both FTSE and IXIC series. These all indicate that the introduction of the attention mechanism does not improve the prediction accuracy in terms of direction.

3) The prediction precision of the proposed VMD-AttGRU model appears to be significantly higher than other compared models except for  $D_{stat}$ . For FTSE and IXIC data, the values of  $D_{stat}$  for VMD-GRU are both the largest, reaching 98.19%, while those for VMD-AttGRU are 98.04 and 94.88%, respectively, which are 0.15 and 3.31% lower than the largest predicted by VMD-GRU.

4) The processing time of hybrid models is significantly longer than that of single models, meaning that the process of establishing and training the forecasting models for each IMF takes longer time. In the comparison of AttGRU with GRU as well as VMD-AttGRU with VMD-GRU, the introduction of attention mechanism layer also leads to a longer processing time. Compared with the LSTM, the processing time of GRU is relative shorter for both FTSE and IXIC, indicating the processing speed by the gates of each hidden layer unit in GRU is slower than that in LSTM.

In brief, following the “Divide-and-Conquer” principle, on the one hand, the proposed hybrid model VMD-AttGRU can improve the forecasting accuracy in terms of error-type performance measures. On the other hand, the introduction of attention mechanism weakens the correctness of predicted direction. Moreover, the “Decomposition-and-Ensemble” framework of the forecasting model inevitably causes greater data processing, which leads to a higher time cost while improving the forecasting quality.



**Figure 8.** The  $D_{stat}$  values of IMFs predicted by AttGRU and GRU for hybrid model.

## 6. Conclusions

A hybrid model VMD-AttGRU is proposed in this study to forecast the stock price indices of FTSE and IXIC. Since the price series is non-stationary and non-linear, the VMD approach is applied to weaken the adverse effect of too much noise in prediction. Moreover, considering that not all elements in the input series contribute equally to the forecasting tasks, the attention mechanism is utilized to assign weights to different input elements for the GRU network and achieves a more

accurate forecasting result. Compared with single models (LSTM, GRU, and AttGRU) and a hybrid model (VMD-GRU), the proposed VMD-AttGRU model exhibits superiority in improving forecasting accuracy of stock price indices after analyzing its performance (MAE, RMSE, MAPE, and TIC) together with trend-type performance ( $D_{stat}$ ). The proposed VMD-AttGRU model can provide an effective paradigm for the prediction of financial time series, which could also be applied to predicting time series in other fields.

## Acknowledgments

The work was partially supported by the Humanities and Social Sciences Foundation of Ministry of Education of China (No. 18YJCZH134) and the Fundamental Research Funds for the Central Universities (No. FRF-BR-18-001B).

## Conflict of interests

The authors declare there is no conflict of interests.

## References

1. C. Zhang, H. Pan, Y. Ma, X. Huang, Analysis of Asia Pacific stock markets with a novel multiscale model, *Phys. A*, **534** (2019), 120939.
2. A. L. D. Loureiro, V. L. Miguéis, L. F. M. da Silva, Exploring the use of deep neural networks for sales forecasting in fashion retail, *Decis. Support Syst.*, **114** (2018), 81–93.
3. J. Wang, J. Wang, Forecasting stock market indexes using principle component analysis and stochastic time effective neural networks, *Neurocomputing*, **156** (2015), 68–78.
4. Y. Xu, S. B. Cohen, *Stock movement prediction from tweets and historical prices*, In Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics, 2018.
5. D.P. Mandic, J.A. Chambers, Exploiting inherent relationships in RNN architectures, *Neural Networks*, **12** (1999), 1341–1345.
6. T. Deng, X. He, Z. Zeng, Recurrent neural network for combined economic and emission dispatch, *Applied Intelligence*, *Appl. Intell.*, **48** (2018), 2180–2198.
7. S. Hochreiter, J. Schmidhuber, Long short-term memory, *Neural Comput.*, **9** (1997), 1735–1780.
8. K. Wang, X. Qi, H. Liu, Photovoltaic power forecasting based LSTM-Convolutional Network, *Energy*, **189** (2019), 116225.
9. Z. Karevan, J. A. K. Suykens, Transductive LSTM for time-series prediction: An application to weather forecasting, *Neural Networks*, **125** (2020), 1–9.
10. B. Zhao, Z. P. Wang, W. J. Ji, X. Gao, X. B. Li, A Short-term Power Load Forecasting Method Based on Attention Mechanism of CNN-GRU, *Power Syst. Technol.*, **12** (2019).
11. Z. Y. Peng, S. Peng, L. D. Fu, B. C. Lu, J. J. Tang, K. Wang, et al., A novel deep learning ensemble model with data denoising for short-term wind speed forecasting, *Energy Convers. Manage.*, **207** (2020), 112524.
12. W. Y. Wu, W. L. Liao, J. Miao, G. L. Du, Using Gated Recurrent Unit Network to Forecast Short-Term Load Considering Impact of Electricity Price, *Energy Procedia*, **158** (2019) 3369–3374.

13. J. Zhang, D. Li, Y. Hao, Z. Tan, A hybrid model using signal processing technology, econometric models and neural network for carbon spot price forecasting, *J. Cleaner Prod.*, **204** (2018), 958–964.
14. J. Wang, L. Y. Tang, Y. Y. Luo, P. Ge, A weighted EMD-based prediction model based on TOPSIS and feed forward neural network for noised time series, *Knowl. Based Syst.*, **132** (2017), 167–178.
15. J. Cao, Z. Li, J. Li, Financial time series forecasting model based on CEEMDAN and LSTM, *Phys. A*, **519** (2019), 127–139.
16. K. Dragomiretskiy, D. Zosso, Variational mode decomposition, *IEEE Trans. Signal Process.*, **62** (2014), 531–544.
17. S. Lahmiri, Intraday stock price forecasting based on variational mode decomposition, *J. Comput. Sci.*, **12** (2016), 23–27.
18. S. Lahmiri, A variational mode decomposition approach for analysis and forecasting of economic and financial time series, *Expert Syst. Appl.*, **55** (2016), 268–273.
19. S. Lahmiri, Comparing variational and empirical mode decomposition in forecasting day-ahead energy prices, *IEEE Syst. J.*, **11** (2015), 1907–1910.
20. Q. Zhu, F. Zhang, S. Liu, Y. Wu, L. Wang, A hybrid VMD–BiGRU model for rubber futures time series forecasting, *Appl. Soft Comput.*, **84** (2019), 105739.
21. C. Li, G. Tang, X. Xue, A. Saeed, X. Hu, Short-term wind speed interval prediction based on ensemble GRU model, *IEEE Trans. Sustainable Energy*, **11** (2020), 1370–1380.
22. R. Wang, C. Li, W. Fu, G. Tang, Deep learning method based on gated recurrent unit and variational mode decomposition for short-term wind power interval prediction, *IEEE Trans. Neural Networks Learn. Syst.*, **31** (2019), 3814–3827.
23. S. Boyd, N. Parikh, E. Chu, B. Peleato, J. Eckstein, *Distributed Optimization and Statistical Learning via the Alternating Direction Method of Multipliers*, Now Foundations and Trends, 2011.
24. Y. Liu, C. Yang, K. Huang, W. Cui, Non-ferrous metals price forecasting based on variational mode decomposition and LSTM network, *Knowl. Based Syst.*, **188** (2020), 105006.
25. J. W. E, J. M. Ye, L. L. He, H. H. Jin, Energy price prediction based on independent component analysis and gated recurrent unit neural network, *Energy*, **189** (2019), 116278.
26. S. Chen, L. Ge, Exploring the attention mechanism in LSTM-based Hong Kong stock price movement prediction, *Quant. Finance*, **19** (2019), 1507–1515.
27. R. Desimon, J. Duncan, Neural mechanisms of selective visual attention, *Annu. Rev. Neurosci.*, **18** (1995), 193–222.
28. M. T. Luong, H. Pham, C. D. Manning, Effective approaches to attention-based neural machine translation, *arXiv:1508.04025*.
29. L. Li, S. Tang, Y. Zhang, L. Deng, Q. Tian, GLA: global-local attention for image description, *IEEE Trans. Multimedia*, **20** (2017), 726–737.
30. S. Wang, X. Wang, S. Wang, D. Wang, Bi-directional long short-term memory method based on attention mechanism and rolling update for short-term load forecasting, *Int. J. Electric. Power Energy Syst.*, **109** (2019), 470–479.

