Received: 21 October 2019 | Revised: 19 January 2020 | Accepted: 20 March 2020

DOI: 10.1002/for.2682

RESEARCH ARTICLE

WILEY

Stock index prediction based on wavelet transform and FCD-MLGRU

Xiaojun Li | **Pan Tang**

School of Economics and Management, Southeast University, Nanjing, Jiangsu, China

Correspondence
Pan Tang, School of Economics and Management, Southeast University, Nanjing, Jiangsu 211189, China,
Email: pantang@seu.edu.cn

Abstract

With the development of artificial intelligence, deep learning is widely used in the field of nonlinear time series forecasting. It is proved in practice that deep learning models have higher forecasting accuracy compared with traditional linear econometric models and machine learning models. With the purpose of further improving forecasting accuracy of financial time series, we propose the WT-FCD-MLGRU model, which is the combination of wavelet transform, filter cycle decomposition and multilag neural networks. Four major stock indices are chosen to test the forecasting performance among traditional econometric model, machine learning model and deep learning models. According to the result of empirical analysis, deep learning models perform better than traditional econometric model such as autoregressive integrated moving average and improved machine learning model SVR. Besides, our proposed model has the minimum forecasting error in stock index prediction.

KEYWORDS

ARIMA, Deep learning, Filter cycle decomposition, Stock index prediction, Wavelet transform

1 | INTRODUCTION

Financial time series forecasting has always been a research focus in finance (Aznarte et al., 2012; Nayak et al., 2018; Podsiadlo & Rybinski, 2016; Pradeepkumar & Ravi, 2017). Autoregressive integrated moving average model (ARIMA) was originally the typical econometric model to forecast time series (Cholette, 1982; Joo & Duk Bin, 1997). However, it has low prediction accuracy for nonlinear time series forecasting. With the rapid development of financial technology (Fintech) (Chen et al., 2019; Itay et al., 2019) these years, interdisciplinary research on the combination of finance and artificial intelligence has been the most active field of financial study. In the aspect of nonlinear financial time series, neural networks (Sagheer & Kotb, 2019; Zhang et al., 2017; Zhang et al., 2019) are in a dominant position in nonlinear financial time series forecasting and relevant studies emerge quickly. For example, Wang et al. (2015) propose a backpropagation neural network with adaptive differential evolution for time series forecasting. Svetlana and Ioannis (2019) propose an ensemble of long short-term memory (LSTM) neural networks for intraday stock predictions.

Wavelet transform (WT;Lee & Shen, 2013; Li et al., 2019) is an ideal tool for time–frequency analysis and processing. In the aspect of financial time series, it can be applied to reduce the impact of noise in forecasting. Moreover, it has been proved from empirical result that FCD-MLGRU (Zhang et al., 2017) performs better than other related deep learning methods. Therefore, as proposed in this study, the WT-FCD-MLGRU model, which is the combination of wavelet transform, filter cycle decomposition and multilag GRU, can theoretically improve the forecasting accuracy of financial time series.

The remainder of this paper is organized as follows: Section 2 introduces the methodology used in the empirical analysis, including wavelet transform (WT), filter cycle decomposition (FCD), variable-length sampling and multilag ensemble forecast methods, gated recurrent unit (GRU), and evaluation metrics for forecasting errors. Section 3 is empirical analysis. Firstly, we choose four major stock indices in the global stock market to test forecasting performance of different models. The chosen stock indices are Standard & Poor's 500 index (S&P 500), NASDAQ Composite (IXIC), Dow Jones Industrial Average (DJI), and the Shanghai Stock Exchange Composite Index (SSE). Secondly, in the aspect of financial time series forecasting, we comprehensively compare and test forecasting accuracy of different models including econometric models such as ARIMA, improved machine learning model like multilag SVR, and our proposed model WT-FCD-MLGRU. The best model for financial time series forecasting is determined by four typical error evaluation metrics. Finally, it is proved from empirical results that WT-FCD-MLGRU has the minimum forecasting error. Section 4 concludes.

2 | METHODOLOGY

2.1 | Multilevel wavelet decomposition and denoising algorithm

Wavelet transform (WT) has recently become a popular method when it comes to the decomposition and denoising of signals. It has been broadly divided into three classes: continuous wavelet transforms (CWT), discrete wavelet transforms (DWT), and multiresolution-based discrete wavelet transforms. Only DWT is introduced because this method is suitable for financial time series. DWT is any wavelet transform for which the wavelets are discretely sampled. Compared with Fourier transforms, one of its key advantages is the temporal resolution, because it can capture both frequency and location information. Let us take one level of DWT as an example. The DWT of a signal x is calculated by passing it through a series of filters. Firstly, the samples are passed through a low-pass filter with impulse response g, resulting in a convolution of the two, which is given by

$$y[n] = (x * g)[n] = \sum_{k=-\infty}^{\infty} x[k]g[n-k]. \quad (1)$$

Secondly, the signal x is also decomposed simultaneously using a high-pass filter h. The outputs are composed of two parts: detail coefficients and approximation coefficients. Detail coefficients are given from the high-pass filter and approximation coefficients are given from the low-pass filter. The two filters, which are known as a quadrature mirror filter, are related to each other.

However, since half the frequencies of the signal x have been removed, half of the samples can be discarded according to Nyquist's rule. After that, the filter output of the low-pass filter g is subsampled by 2 and further processed by passing it again through a new low-pass filter g and a high-pass filter h with half the cut-off frequency of the previous one, which is given by

$$y_{\text{low}}[n] = \sum_{k=-\infty}^{\infty} x[k]g[2n-k], \quad (2)$$

$$y_{\text{high}}[n] = \sum_{k=-\infty}^{\infty} x[k]h[2n-k]. \quad (3)$$

According to $(y{\downarrow}k)[n] = y[kn]$, the above summation now can be more concisely written by

$$y_{\text{low}} = (x * g) \downarrow 2, \quad (4)$$

$$y_{\text{high}} = (x * h) \downarrow 2, \quad (5)$$

where $\downarrow$ denotes the subsampling operator.

The level 4 decomposition diagram of DWT, which is used in empirical analysis, is illustrated in Figure 1. cA_n denotes the approximation coefficients and cD_n denotes the detail coefficients.

A wavelet threshold denoising algorithm was proposed to eliminate the effect of noise. Fixed-form threshold ($\sqrt{2 * \ln[\text{length}(x)]}$, where length($x$) denotes the length of signal x), commonly used in the wavelet threshold denoising algorithm, is chosen to eliminate the noise of financial time series.

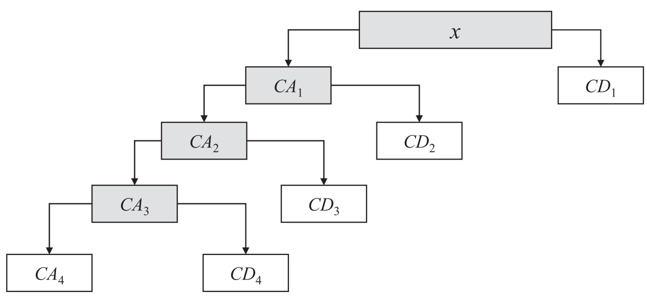


FIGURE 1 Level 4 decomposition diagram of DWT

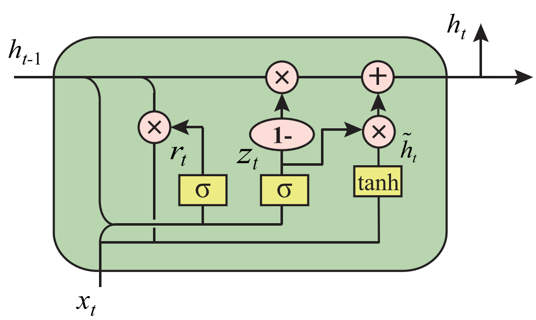


FIGURE 2 Structure diagram of GRU [Colour figure can be viewed at wileyonlinelibrary.com]

2.2 | Forecasting model

The forecasting model WT-FCD-MLGRU is the improved GRU model, which is the combination of wavelet transform, filter cycle decomposition, and multilag neural networks.

2.2.1 | Gated recurrent unit (GRU)

Considering that there are various neural networks in the field of artificial intelligence, some widely used neural networks are chosen to compare their forecasting performance in forecasting stock index. According to the empirical result, our improved model based on gated

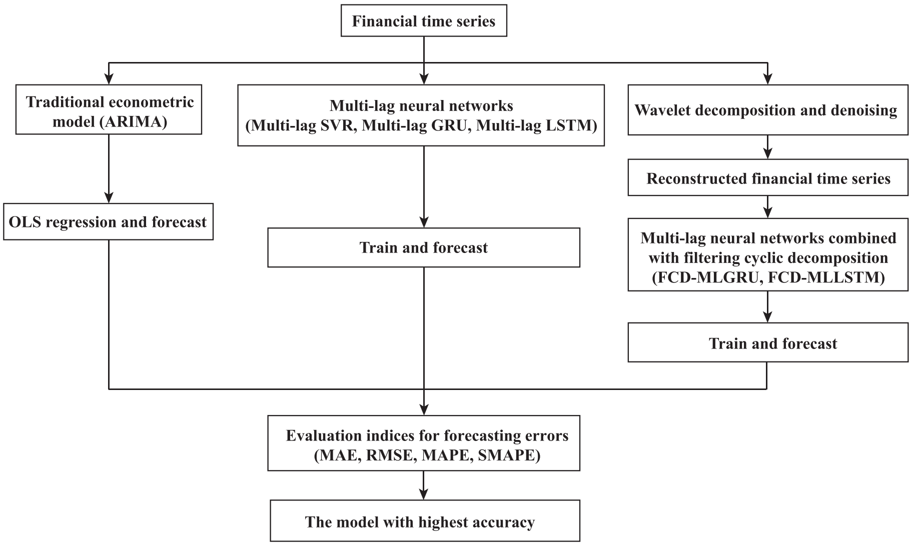


FIGURE 3 Flow chart of empirical analysis

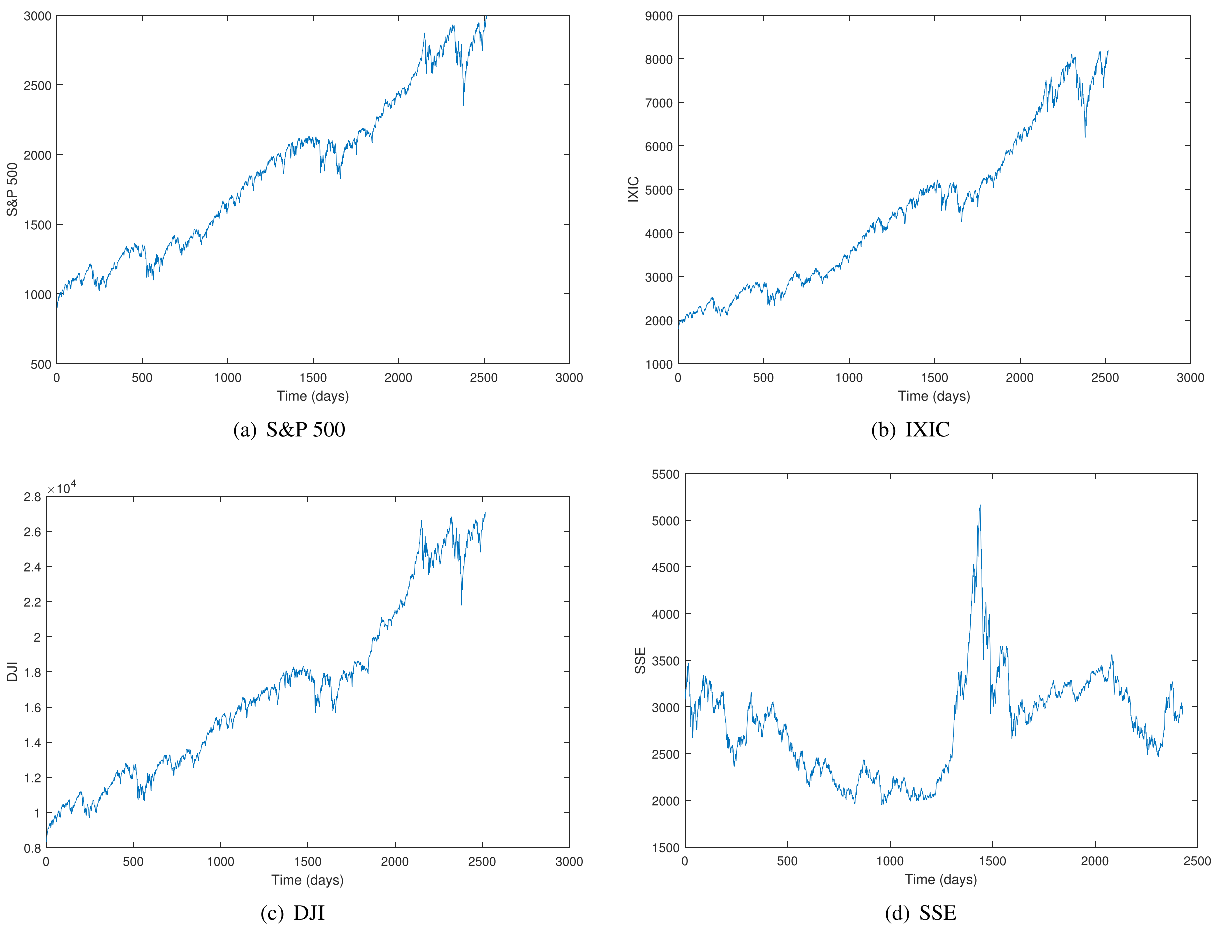


FIGURE 4 Original time series of four stock indices' daily price. [Colour figure can be viewed at wileyonlinelibrary.com]

recurrent unit (GRU) has the highest forecasting accuracy. GRU is a variation of recurrent neural network (RNN). It has two gate structures: update gate $\boldsymbol{z_t}$ and reset gate $\boldsymbol{r_t}$. The function of $\boldsymbol{z_t}$ is to resolve how much the neural unit updates, and the function of $\boldsymbol{r_t}$ is to resolve how much previously state the neural unit forgets. The mathematical expressions of $\boldsymbol{z_t}$ and $\boldsymbol{r_t}$ are given by

$$\boldsymbol{z_t} = \tanh(\boldsymbol{W_z} \cdot [\boldsymbol{h_{t-1}}, \boldsymbol{x_t}]), \tag{6}$$

$$\boldsymbol{r_t} = \tanh(\boldsymbol{W_r} \cdot [\boldsymbol{h_{t-1}}, \boldsymbol{x_t}]), \tag{7}$$

where $\boldsymbol{W}$ denotes the weight, $\boldsymbol{h_{t-1}}$ denotes the last output, $\boldsymbol{x_t}$ denotes input and "·" denotes element-wise product.

The neural units memory as $\tilde{\boldsymbol{h}}_t$ and expose state as output $\boldsymbol{h_t}$ are then calculated by

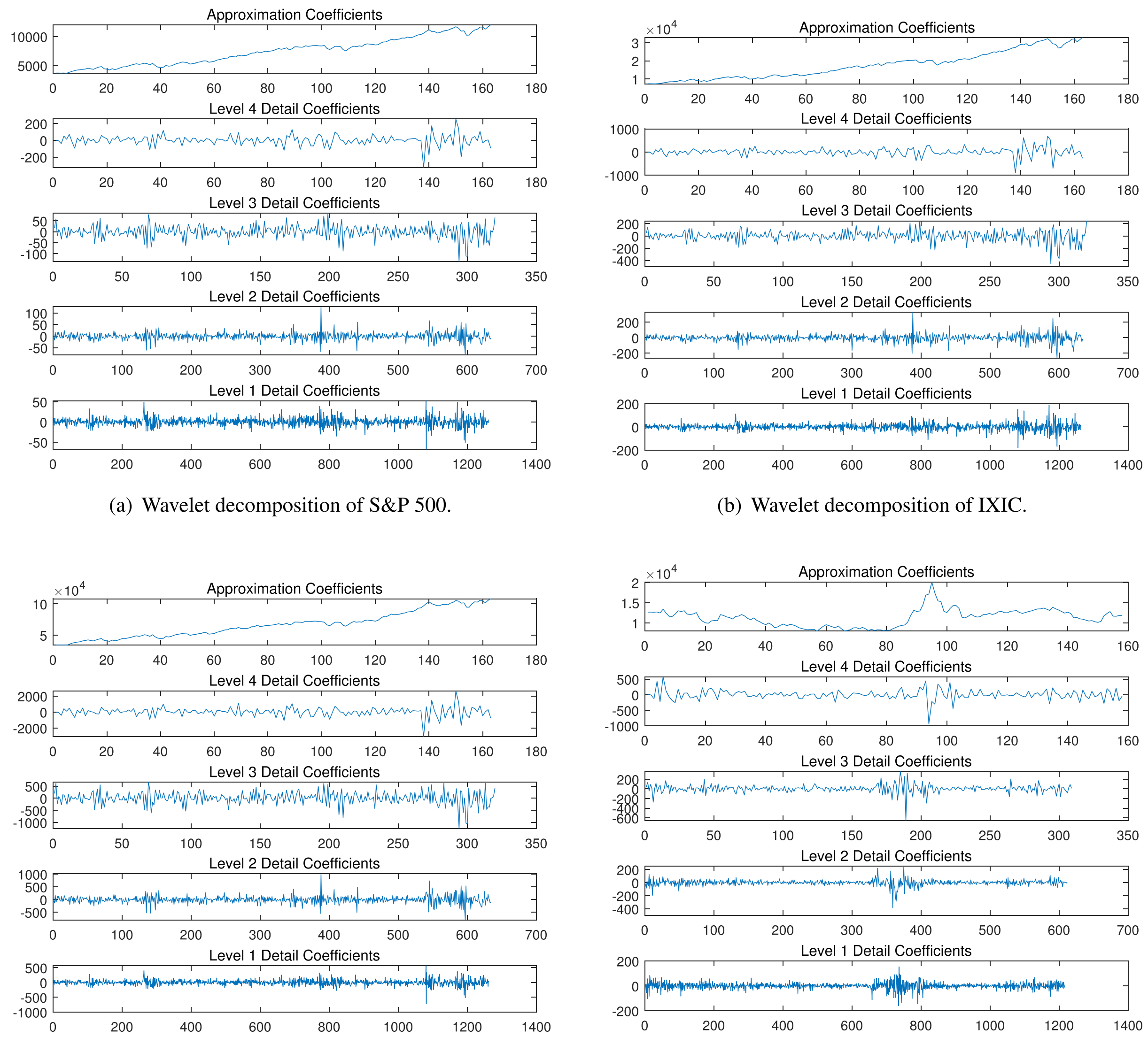
$$\tilde{\boldsymbol{h}}_t = \tanh(\boldsymbol{W} \cdot [\boldsymbol{r_t} * \boldsymbol{h_{t-1}}, \boldsymbol{x_t}]), \tag{8}$$

$$\boldsymbol{h_t} = (\boldsymbol{1} - \boldsymbol{z_t}) * \boldsymbol{h_{t-1}} + \boldsymbol{z_t} * \tilde{\boldsymbol{h}}_t. \tag{9}$$

The structure diagram of GRU is shown in Figure 2.

TABLE 1 Descriptive statistics of four stock indices

Index	Count	Mean	Max	Min	SD	Skewness	Kurtosis
S&P 500	2,517	1,868.24	2,999.91	901.05	569.02	0.21	-1.13
IXIC	2,517	4,448.78	8,202.53	1,793.20	1,784.52	0.47	-0.93
DJI	2,517	16,776.80	27,088.08	8,331.68	4,940.16	0.45	-0.84
SSE	2,429	2,808.17	5,166.35	1,950.01	535.68	0.71	1.34



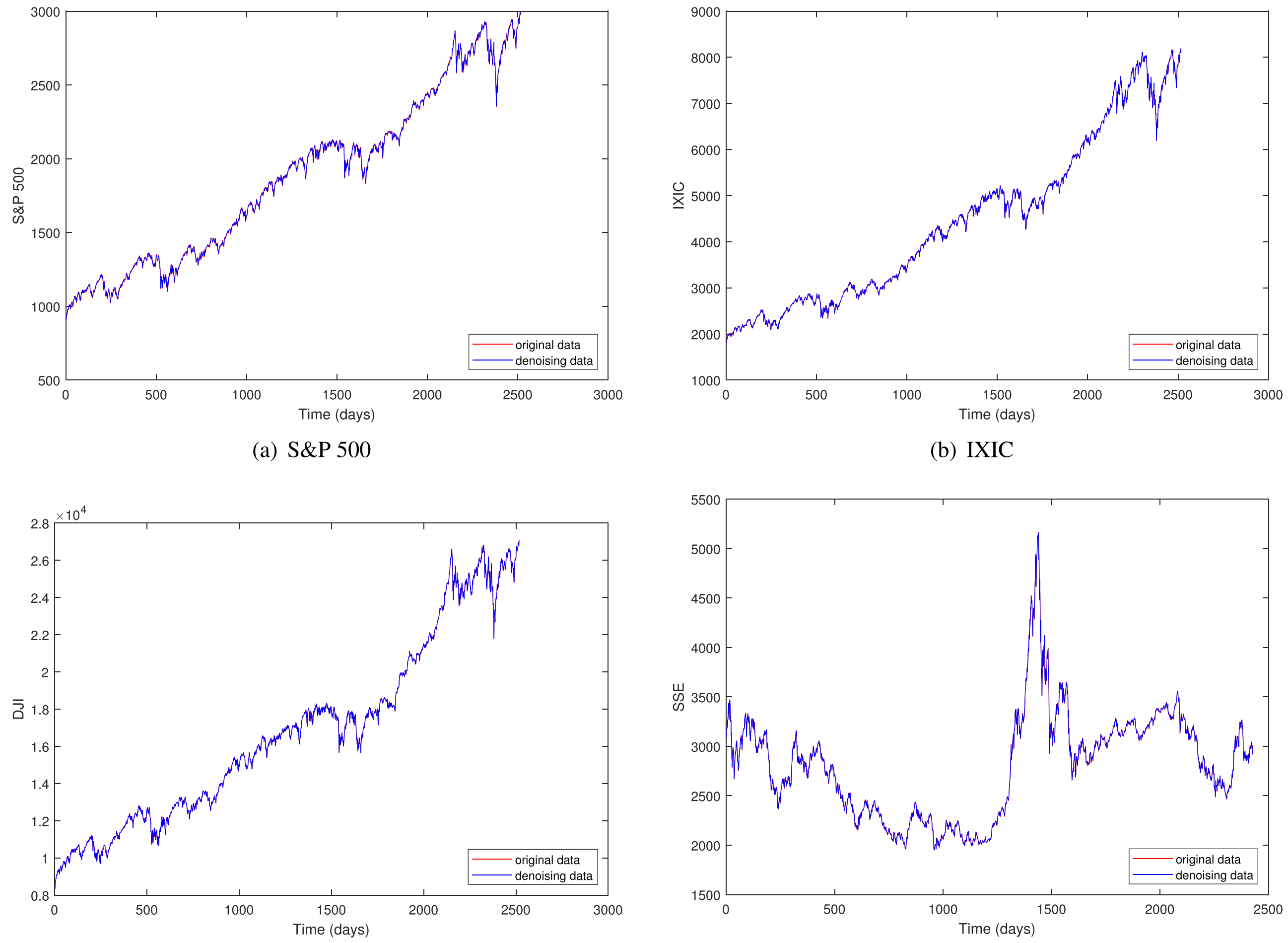
(a) Wavelet decomposition of S&P 500.

(b) Wavelet decomposition of IXIC.

(c) Wavelet decomposition of DJI.

(d) Wavelet decomposition of SSE.

FIGURE 5 Wavelet decomposition of four stock indices [Colour figure can be viewed at wileyonlinelibrary.com]



(a) S&P 500 (b) IXIC (c) DJI (d) SSE

FIGURE 6 Difference between original data and denoising data [Colour figure can be viewed at wileyonlinelibrary.com]

TABLE 2 SNR and MSE of denoising data

Index	SNR	MSE
S&P 500	127.1263	11.5746
IXIC	125.6840	13.5050
DJI	148.0353	14.3235
SSE	127.0495	13.3014

2.2.2 | Filter cycle decomposition

According to the theory of time series analysis, a time series $Y(t)$ can be decomposed as

$$Y(t) = T(t) + C(t) + R(t), \tag{10}$$

where $T(t)$ denotes the trend time series, $C(t)$ denotes the cyclic time series, and $R(t)$ denotes the residual time series.

Based on the above theory, the filter cycle decomposition (FCD) method is defined as follows:

1. Choose a fixed cyclic period a. Use $2 * a$ (a is even) or a (a is odd) window size to do the moving average (MA) to estimate the trend subseries T.
2. Remove the trend information $Y - T$.
3. Calculate the average of each moment in the cyclic period as cyclic a information C.
4. Gain residual information $R = Y - T - C$.

2.2.3 | Variable-length sampling and multilag ensemble forecasting method

The variable-length time lag sampling method is applied to further improve the forecasting accuracy of deep neural networks. At the time of training, the value of next time point x_{t+1} is forecast as ground-truth at first. Then, a minimum length $L_{\min}$ and a maximum length $L_{\max}$ were specified. Next, sequences are cut in all possible lengths within $L_{\min} \sim L_{\max}$ prior to this point as regressor input, which can be described in detail as $[x_t, x_{t-1}, \ldots, x_{t-L_{\min}}]$, $[x_t, x_{t-1}, \ldots, x_{t-L_{\min}+1}]$, $[x_t, x_{t-1}, \ldots, x_{t-L_{\max}}]$. In the period of forecasting, each neural network will obtain different forecasting results in different time lags, so we take their average as each neural network's final forecasting results.

The variable-length time lag sampling method can effectively reduce the instability of single time lag forecasting results. It can even maintain prediction ability in the case

TABLE 3 The forecasting errors (MAE, RMSE, MAPE and SMAPE) of four stock indices

Index	Model	Lag window	Epoch	MAE	RMSE	MAPE	SMAPE
S&P 500	ARIMA(2, 1, 2)	—	—	91.2902	99.8437	3.1102	3.1671
	Multilag SVR	24	—	51.0570	54.3884	1.7525	1.7652
	Multilag LSTM	24	200	29.2241	34.4745	1.0055	1.0122
	Multilag GRU	24	200	16.8599	20.8488	0.5863	0.5856
	WT-FCD-MLLSTM	24	40	17.1913	22.3404	0.6052	0.6027
	WT-FCD-MLGRU	**24**	**40**	**11.6431**	**14.0426**	**0.4065**	**0.4066**
IXIC	ARIMA(2, 1, 1)	—	—	207.2802	231.9820	2.5987	2.6362
	Multilag SVR	24	—	157.3930	173.4611	1.9912	2.0045
	Multilag LSTM	24	200	68.0935	84.0275	0.8733	0.8711
	Multilag GRU	24	200	78.0267	99.2380	0.9943	0.9908
	WT-FCD-MLLSTM	24	40	66.6168	77.5410	0.8528	0.8554
	WT-FCD-MLGRU	**24**	**40**	**50.3540**	**60.9144**	**0.6470**	**0.6474**
DJI	ARIMA(3, 1, 3)	—	—	831.9505	901.0102	3.1391	3.1960
	Multilag SVR	24	—	455.1997	481.1955	1.7309	1.7429
	Multilag LSTM	24	200	170.7660	205.6517	0.6559	0.6562
	Multilag GRU	24	200	127.4530	177.9483	0.4888	0.4902
	WT-FCD-MLLSTM	24	40	151.1917	184.4279	0.5863	0.5861
	WT-FCD-MLGRU	**24**	**40**	**110.6576**	**142.6386**	**0.4292**	**0.4289**
SSE	ARIMA(3, 1, 2)	—	—	62.7791	78.3785	2.1083	2.1402
	Multilag SVR	24	—	35.7007	44.6939	1.2105	1.2149
	Multilag LSTM	24	200	23.6867	32.9050	0.8030	0.8055
	Multilag GRU	24	200	22.6602	31.3590	0.7698	0.7708
	WT-FCD-MLLSTM	24	40	17.1628	20.1477	0.5871	0.5865
	WT-FCD-MLGRU	**24**	**40**	**12.7347**	**16.9160**	**0.4346**	**0.4347**

when time lag drifts in the datastream and improve the generalization ability of the forecasting model.

2.3 | Evaluation metrics for forecasting errors

Four typical evaluation metrics for forecasting errors are applied to measure forecasting accuracy of each model, including mean absolute error (MAE), root mean square error (RMSE), mean absolute percentage error (MAPE), and symmetric mean absolute percentage error (SMAPE). They are calculated as follows:

$$\mathrm{MAE} = \frac{1}{n}\sum_{t=1}^{n} |y_t - \hat{y}_t|, \tag{11}$$

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{t=1}^{n}(y_t - \hat{y}_t)^2}, \tag{12}$$

$$\mathrm{MAPE} = 100 \times \frac{1}{n}\sum_{t=1}^{n}\left|\frac{y_t - \hat{y}_t}{y_t}\right|, \tag{13}$$

$$\mathrm{SMAPE} = 100 \times \frac{1}{n}\sum_{t=1}^{n}\frac{|\hat{y}_t - y_t|}{(|\hat{y}_t| + |y_t|)/2}, \tag{14}$$

where n denotes the total number of the test samples, y_t denotes the actual value at time t and $\hat{y}_t$ denotes the predicted value at time t.

3 | EMPIRICAL ANALYSIS

3.1 | Flow chart of empirical analysis

The empirical analysis for financial time series forecasting is composed of four parts. Firstly, we make the prediction based on traditional econometric model ARIMA by ordinary least square (OLS) regression and take its forecasting result as the benchmark. Secondly, multilag neural networks like multilag GRU and multilag LSTM are applied to forecast the stock indices. Thirdly, wavelet transform is used to decompose and denoise the stock indices. FCD-MLGRU and FCD-MLLSTM, which are a combination of multilag neural networks and filter cycle decomposition, are then chosen to train and forecast the reconstructed stock indices. Finally, the model with highest forecasting accuracy is selected by evaluation metrics for forecasting errors. A flow chart of empirical analysis is shown in Figure 3.

3.2 | Statistical analysis of stock indices

Four major stock indices—Standard & Poor's 500 index (S&P 500), NASDAQ Composite (IXIC), Dow Jones Industrial Average (DJI), and the Shanghai Stock Exchange Composite Index (SSE)—are selected to implement empirical analysis. The original data consist of the daily closing prices of these indices, all of which are

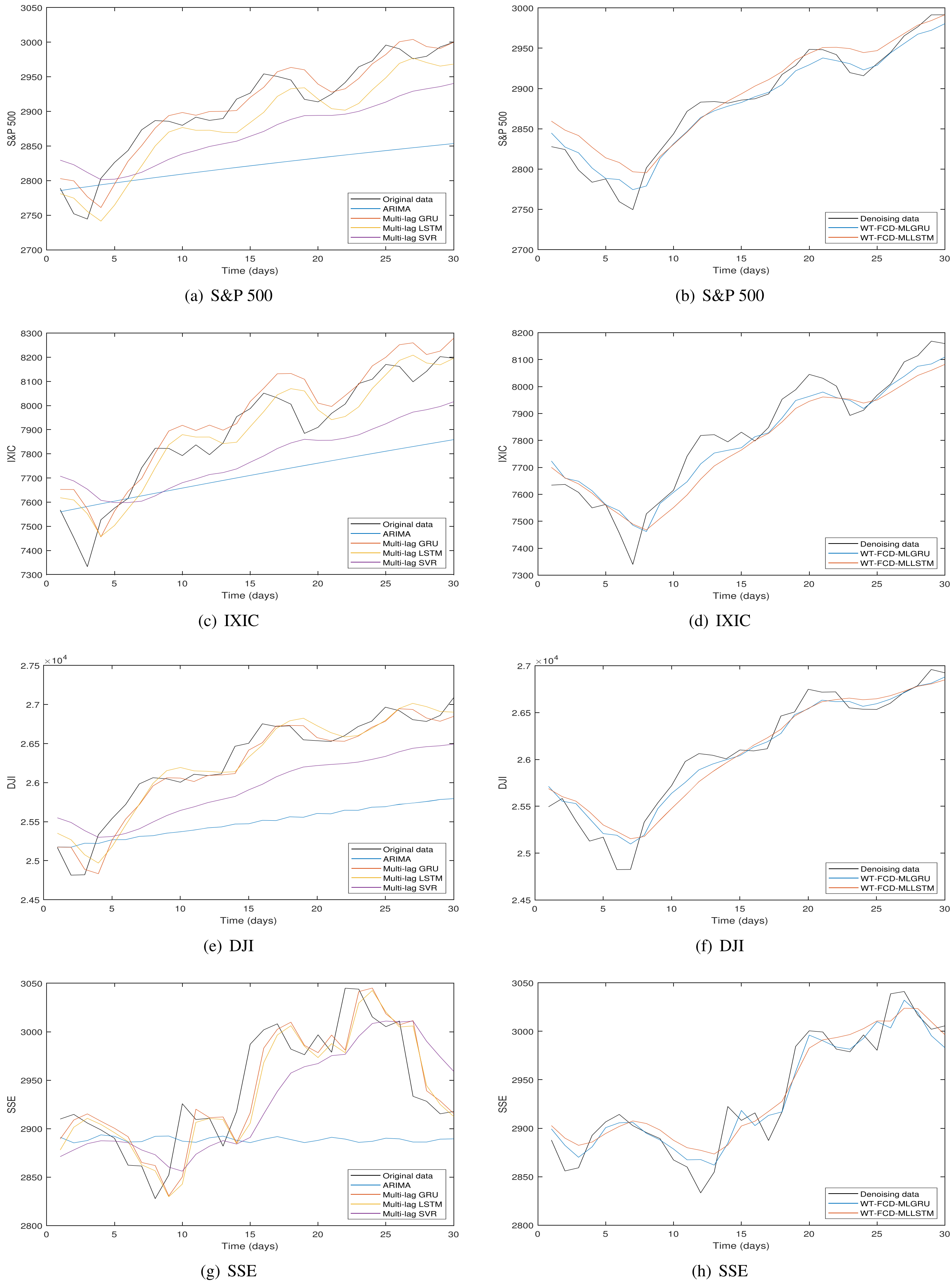


FIGURE 7 Difference between true value and predicted value [Colour figure can be viewed at wileyonlinelibrary.com]

downloaded from Yahoo Finance (https://finance.yahoo.com). The time range of data is from July 12, 2009, to July 12, 2019. The original daily closing price time series of S&P 500, IXIC, DJI, and SSE are illustrated in Figure 4 and descriptive statistics are shown in Table 1.

3.3 | Wavelet transform

3.3.1 | Wavelet decomposition process

Wavelet Daubechies 4 (db4) is selected as wavelet base. According to the theory of multiresolution, a higher decomposition level will enhance the denoising effect because more low-frequency components are removed with the increase in decomposition level. However, the side effect is that a higher decomposition level will lead to the distortion increasing. Therefore, level 4 is chosen as the decomposition level according to the feature of financial time series. Wavelet decomposition of stock indices is shown in Figure 5.

3.3.2 | Wavelet denoising process

After wavelet decomposition, fixed-form threshold $\sqrt{2 * \ln[\text{length}(x)]}$, where length($x$) denotes the length of stock index time series, is used to eliminate the noise of detail coefficients because a large amount of noise lies in detail coefficients. The time series are then reconstructed by denoising detail coefficients and approximation coefficients. Figure 6 shows the difference between the original data and the denoising data.

As is known, signal-to-noise ratio (SNR) and mean square error (MSE) are broadly used to measure the quality of denoising data. it can be seen from Table 2 that the noise of stock indices is effectively eliminated.

3.4 | Forecasting results of different models

Each stock index's data are divided into two parts: the last 30 days' price data are the test data and the rest are the training data. As illustrated in the flow chart shown in Figure 3, the traditional ARIMA econometric model is first used to forecast the last 30 days' prices. p and q in ARIMA (p, d, q) are selected by Akaike information criterion (AIC). d in ARIMA (p, d, q) is chosen by unit root test. Then, multilag neural networks, including multilag SVR, multilag GRU and multilag LSTM, are used to train and forecast the last 30 days' prices of these stock indices. In order to compare the forecasting accuracy of different models, the lag window and epoch of each neural network are set to 24 and 200, respectively. Next, we apply WT-FCD-MLGRU and WT-FCD-MLLSTM, which are both the combination of wavelet transform and multilag neural networks with filtering cycle decomposition, to predict stock indices. Lag window and epoch are set to 24 and 40, respectively. Finally, four different measuring indices for forecasting errors are used to evaluate each model's forecasting performance. Table 3 shows the forecasting errors (MAE, RMSE, MAPE and SMAPE) and Figure 7 intuitively shows the difference between the true value and predicted value.

As can be seen from Figure 7 and Table 3, the traditional ARIMA econometric models suitable for linear forecasting are not fitted to forecast nonlinear financial time series. Multilag SVR, which is an improved model based on the SVR machine learning model, also has a higher forecasting error compared with improved deep learning models. Besides, the WT-FCD-MLGRU model proposed in this work has the highest accuracy among these models in forecasting 30 days' daily prices of S&P 500, IXIC, DJI and SSE. Its forecasting error is much lower, especially for those stock indices with smaller values.

4 | CONCLUSION

In this study, the WT-FCD-MLGRU model is proposed for financial time series forecasting and subsequently applied to forecast daily closing prices of four major stock indices. Compared with the traditional ARIMA econometric model, the improved SVR machine learning model, and other multilag neural networks, our proposed model, which is the combination of wavelet transform, filter cycle decomposition, and multilag GRU, has the highest accuracy in forecasting stock index. It is especially applicable to predicting those stock indices with smaller values. More importantly, it is proved in empirical analysis that when facing financial big data deep learning models would be the first choice for their higher forecasting accuracy, no need for feature engineering, and stronger ability to adapt than linear econometric models or traditional machine learning models.

ACKNOWLEDGMENTS

The authors gratefully acknowledge financial support from the National Social Science Fund of China (No. 19CJL028).

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable request.

REFERENCES

Aznarte, J. Z., Alcalá-Fdez, J., Arauzo-Azofra, A., & Benítez, J. M. (2012). Financial time series forecasting with a bio-inspired fuzzy model. *Expert Systems with Applications*, *39*(16), 12302–12309.

Chen, M. A., Wu, Q., & Yang, B. (2019). How valuable is FinTech innovation. *Review of Financial Studies*, *32*(5), 2062–2106.

Cholette, P. A. (1982). Prior information and ARIMA forecasting. *Journal of Forecasting*, *1*(4), 375–383.

Itay, G., Wei, J., & Andrew, K. G. (2019). To FinTech and beyond. *Review of Financial Studies*, *32*(5), 1647–1661.

Joo, Y. J., & Duk Bin, J. (1997). State space trend-cycle decomposition of the ARIMA(1,1,1) process. *Journal of Forecasting*, *16*(6), 411–424.

Lee, T., & Shen, H. (2013). Efficient local statistical analysis via integral histograms with discrete wavelet transform. *IEEE Transactions on Visualization and Computer Graphics*, *19*(12), 2693–2702.

Li, H., Fan, P., Xia, H., & Song, S. (2019). Quantum multi-level wavelet transforms. *Information Sciences*, *504*, 113–135.

Nayak, S. C., Misra, B. B., & Behera, H. S. (2018). Artificial chemical reaction optimization based neural net for virtual data position exploration for efficient financial time series forecasting. *Ain Shams Engineering Journal*, *9*(4), 1731–1744.

Podsiadlo, M., & Rybinski, H. (2016). Financial time series forecasting using rough sets with time-weighted rule voting. *Expert Systems with Applications*, *66*, 219–233.

Pradeepkumar, D., & Ravi, V. (2017). Forecasting financial time series volatility using particle swarm optimization trained quantile regression neural network. *Applied Soft Computing*, *58*, 35–52.

Sagheer, A., & Kotb, M. (2019). Time series forecasting of petroleum production using deep LSTM recurrent networks. *Neurocomputing*, *323*, 203–213.

Svetlana, B., & Ioannis, T. (2019). An ensemble of LSTM neural networks for high-frequency stock market classification. *Journal of Forecasting*, *38*(6), 600–619.

Wang, L., Zeng, Y., & Chen, T. (2015). Back propagation neural network with adaptive differential evolution algorithm for time series forecasting. *Expert Systems with Applications*, *42*(2), 855–863.

Zhang, X., Shen, F., Zhao, J., & Yang, G. (2017). Time series forecasting using GRU neural network with multi-lag after decomposition. In Liu, D., Xie, S., Li, Y., Zhao, D., & El-Alfy, E. S. (Eds.), *Neural information processing, ICONIP 2017 (pp.523–532). Lecture Notes in Computer Science*, Vol. *10638*.

Zhang, Y., Wang, X., & Tang, H. (2019). An improved Elman neural network with piecewise weighted gradient for time series prediction. *Neurocomputing*, *359*, 199–208.

AUTHOR BIOGRAPHIES

Pan Tang, is currently an Associate Professor at the School of Economics and Management, Southeast University. He received his PhD degree in quantum finance from the National University of Singapore. His research interests are financial artificial intelligence, deep learning in finance, quantitative investment and financial big data analysis.

Xiaojun Li, is a PhD student, studying at the School of Economics and Management, Southeast University. He obtained his Master's and Bachelor's degrees from the School of Finance, Nanjing University of Finance and Economics. His research field is financial big data analysis, including stock price prediction based on artificial intelligence and financial risk management.

How to cite this article: Li X, Tang P. Stock index prediction based on wavelet transform and FCD-MLGRU. *Journal of Forecasting*. 2020;39: 1229–1237. https://doi.org/10.1002/for.2682