Forecasting Method of Stock Market Volatility in Time Series Data Based on Mixed Model of ARIMA and XGBoost

Yan Wang, Yuankai Guo*

College of Computer and Communication, LanZhou University of Technology, Lanzhou 730050, China
* The corresponding author, e-mail: 646288897@qq.com.

Abstract: Stock price forecasting is an important issue and interesting topic in financial markets. Because reasonable and accurate forecasts have the potential to generate high economic benefits, many researchers have been involved in the study of stock price forecasts. In this paper, the DWT-ARIMA-GSXGB hybrid model is proposed. Firstly, the discrete wavelet transform is used to split the data set into approximation and error parts. Then the ARIMA (0, 1, 1), ARIMA (1, 1, 0), ARIMA (2, 1, 1) and ARIMA (3, 1, 0) models respectively process approximate partial data and the improved xgboost model (GSXGB) handles error partial data. Finally, the prediction results are combined using wavelet reconstruction. According to the experimental comparison of 10 stock data sets, it is found that the errors of DWT-ARIMA-GSXGB model are less than the four prediction models of ARIMA, XGBoost, GSXGB and DWT-ARIMA-XGBoost. The simulation results show that the DWT-ARIMA-GSXGB stock price prediction model has good approximation ability and generalization ability, and can fit the stock index opening price well. And the proposed model is considered to greatly improve the predictive performance of a single ARIMA model or a single XGBoost model in predicting stock prices.

Keywords: hybrid model; discrete wavelet transform; ARIMA; XGBoost, grid search; stock price forecast

I. INTRODUCTION

Stock price forecasting is an important issue in financial markets. It is also a classic and interesting topic. Because reasonable and accurate forecasts have the potential to generate high economic benefits, many researchers have been involved in the study of stock price forecasts. As early as 1991, McQueen, Grant used the Markov chain to verify the feasibility of stock price forecasting, thus laying a solid foundation for subsequent research [1]. In the following decades, scholars used some advanced machine learning techniques to predict stock prices. At present, different stock price forecasting techniques have been developed. time series predictions such as autoregressive integrated moving average (ARIMA), autoregressive conditional heteroskedasticity (ARCH) and generalized autoregressive conditional heteroskedasticity (GARCH) [2-4], and machine learning-based techniques such as

Received: Jul. 5, 2019
Revised: Oct. 13, 2019
Editor: Dongliang Xie

This paper proposes a DWT-ARIMA-GSXGB hybrid model, greatly improving the predictive performance of a single ARIMA model or a single XGBoost model in predicting stock prices

neural Network, deep learning, support vector machine (SVM) and decision tree, have been widely used in stock price predictions [5-9]. K Kim used the SVM for the prediction of the stock price index and compared it with back-propagation neural networks and case-based reasoning methods. The results show that SVM is a promising stock market forecasting method [10]. Eric F. Oteng-Abayie used the random walk (RW), GARCH (1, 1), EGARCH (1, 1) and TGARCH (1, 1) models to model the fluctuations (conditional variance) of the Ghana Stock Exchange and prediction. The results show that the GARCH(1, 1) model is superior to other models under the assumption that innovation follows a normal distribution [11]. LA Laboissiere used artificial neural networks (ANNs) to realistically predict the closing price of Brazilian power distribution companies, which had their performances evaluated by means of Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE) and Root Mean Square Error (RMSE) calculations. It shows that artificial neural network has better prediction performance in time series prediction [12]. Ariyo built the ARIMA (2, 1, 0) stock price prediction model and used it in the short-term forecast of stock price. Results obtained revealed that the ARIMA model has a strong potential for short-term prediction [13].

According to the above research, it is not difficult to find that most of the previous studies in time series prediction are the improvement and application of a single model.Because there are linear parts in the stock data, and there are also nonlinear parts, the use of combined models for prediction has become a popular research by many scholars. Since the early work of Bates and Granger, several combinations of predictive structures have been explored [14]. Clemen conducted a comprehensive bibliographic review in the field [15]. Menezes et al. provided good guidance for combined forecasting [16]. They concluded that the problem of combined forecasting was to implement a multi-standard process and judge the attributes of the wrong specification. Lam et al. proposed a target programming model to obtain the optimal weight of the combined prediction model [17]. G. Peter Zhang used the unique advantages of ARIMA and ANN models in linear and nonlinear modeling to construct a stock price prediction model combining ARIMA and ANN [18]. The results show that the combined model can effectively improve the prediction accuracy of a single model. Zhang, Yan proposed a new stock index forecasting method based on wavelet analysis combines the autoregressive integrated moving average (ARIMA) and artificial neural network. The non-stationary share price index series are decomposed and reconstructed into one low frequency signal and several high frequency signals by wavelet; the approximate stationary low frequency signal is predicted using ARIMA forecasting model, and the high frequency signals are forecasted using Elman neural network models; the prediction result of each layer are mixed by the radial basis function(RBF)neural network and the result is the final prediction. The example shows that the combined prediction model has higher prediction accuracy [19]. Shuzhen Shi proposed a stock price forecasting model combining ARMA, BPNN and Markov models [20]. The results show that the ARMA-BPNN model is better than the single ARMA model and BPNN model. Ye et al. presented a stock forecasting model based on wavelet analysis and ARIMA-SVR [21]. The stock price was decomposed into reconstructed part and error part by wavelet decomposition and wavelet reconstruction. Then, the ARIMA model and the SVR model was used to forecast the reconstructed part and the error part respectively, and the final prediction results was combined to obtain the final prediction results. The experimental results show that compared with the single forecasting model, the proposed model is an effective method for forecasting stock price, which greatly improves the accuracy of forecasting.

This study presents a hybrid model of discrete wavelet transform, ARIMA and optimized XGBoost(GSXGB) [22-23] to slove the stock price forecasting problem(DWT-ARI-

MA-GSXGB). The DWT-ARIMA-GSXGB hybrid model is proposed. Firstly, the discrete datalet transform is used to split the data set into approximation and error parts. Then the ARIMA model processes the approximate partial data and the improved xgboost model (GSXGB) handles error partial data. Finally, the prediction results are combined using wavelet reconstruction. According to the experimental comparison of 10 stock data sets, it is found that the errors of DWT-ARIMA-GSXGB model are less than the four prediction models of ARIMA, XGBoost, GSXGB and DWT-ARIMA-XGBoost. The simulation results show that the DWT-ARIMA-GSXGB stock price prediction model has good approximation ability and generalization ability, and can fit the stock index opening price well. And the proposed model is considered to greatly improve the predictive performance of a single ARIMA model or a single XGBoost model in predicting stock prices.

II. HYBRID MODEL IN FORECASTING

2.1 Discrete wavelet transform model

Spectrum analysis based on Fourier transform is the most commonly used tool for frequency domain analysis [24]. According to Fourier theory, a signal can be expressed as a sum of a series of sine and cosine. However, a serious limitation of the Fourier transform is that it does not provide any information about the spectral changes in time. Wavelet transform (WT) [25] is an ideal tool for signal time-frequency analysis and processing, similar to Fourier transform, and is a new transform analysis method. It inherits and develops the idea of short-time Fourier transform localization, and at the same time overcomes the shortcomings of window size not changing with frequency, and can provide a "time-frequency" window that changes with frequency. Through the expansion and translation operations, the signal (function) is gradually multi-scale refined, and finally reaches the high-frequency time subdivision and the low-frequency frequency subdivision. Discrete Wavelet Transform (DWT) automatically adapts to the requirements of time-frequency signal analysis and focuses on any detail of the signal. When analyzing time series data in finance, the original sequence can be divided into low frequency (approximate) and high frequency (error) components while maintaining orthogonality for multiresolution analysis.

The wavelet transform formula is shown in equation (1).

$$\varphi_{a,b}(t) = \frac{1}{\sqrt{a}} \varphi(\frac{t-b}{a}) , \qquad (1)$$

Where a represents the shift factor. b represents the displacement factor. a, b are constants, and $a > 0$. $\varphi_{a,b}(t)$ is the basic function $\varphi(t)$ is first shifted and then scaled. If a, b are constantly changing, a cluster of functions $\varphi_{m,n}(t)$ can be obtained. Given a squared integrable signal $f(t)$, ie $f(t) \in L^2(R)$, the corresponding discrete wavelet transform is as shown in equation (2).

$$W_f(m,n) = \langle f, \varphi_{m,n}(t) \rangle = \int_{-\infty}^{+\infty} f(t) \varphi_{m,n}^*(t) dt, \quad (2)$$

Where <*, *> represents the inner product. * represents the complex conjugate. m, n are the values of the discretization processing of the scale factor a and the displacement factor b.

The stock historical data set is a financial time series data set that can be decomposed into an approximate partial data set and an error partial data set. Since discrete wavelet transform has excellent performance in financial time series data set decomposition, DWT is used to decompose the stock data set and decompose it into approximate part and error part.

2.2 ARIMA

The ARIMA model was pioneered by Box and Jenkins and is one of the most popular prediction methods in time series prediction [26]. ARIMA(p, d, q) is called differential autoregressive moving average model, AR is autoregressive, MA is moving average, p and q are corresponding orders, and d is the number of times the time series becomes stationary. The model is a linear regression model that is used to track linear trends in stationary time series data where future values of time series are generated from linear functions observed in the past. The ARIMA (p, d, q) model essentially performs the d-order differential processing on the non-stationary stock historical data Y_t to obtain a new stable stock history sequence X_t, fits the X_t to the ARMA(p, q) model, and then The original d-time differential restoration can obtain the predicted data of Y_t. Among them, the general expression of ARMA (p, q) is shown in formula (3).

$$\begin{aligned} Y_t &= \varphi_0 + \varphi_1 Y_{t-1} + \ldots + \varphi_p Y_{t-p} \\ &+ \varepsilon_t - \theta_1 \varepsilon_{t-1} - \ldots - \theta_q \varepsilon_{t-q}. \end{aligned} \quad (3)$$

The first half of the equation is the autoregressive part, p is the autoregressive order, and φ_p is the autoregressive coefficient. The second half is the moving average part, q is the moving average order, Θ_q is the moving average coefficient; Y_t is the correlation sequence of the consumed stock data, and ε_t is the random error. Basically, this method has three phases: model identification, parameter estimation and diagnostic checking. For example, the ARIMA(1, 1, 1) model can be represented as follows.

$$Y_t = \varphi_0 + \varphi_1 Y_{t-1} + \varepsilon_t - \theta_1 \varepsilon_{t-1}. \quad (4)$$

The ARIMA model is basically a data-oriented approach that is adapted from the structure of the data themselves. However, any error partial data set limit the ARIMA. Therefore, the proposed hybrid model used the eXtreme Gradient Boosting (XGBoost) to deal with the error partial data.

2.3 eXtreme gradient boosting

eXtreme Gradient Boosting was proposed by Tianqi Chen in 2015 and is one of the boosting algorithms. It is proved in the literature [23] that the XGBoost model has the characteristics of low computational complexity, fast running speed and high accuracy. The idea of the Boosting algorithm is to integrate many weak classifiers into a strong classifier. Because XGBoost is a lifting tree model, it integrates many tree models to form a strong classifier. The tree model used is the CART (Classification and Regression Tree) regression tree model [27].

2.3.1 CART regression tree

The generation of decision trees is the process of recursively constructing a binary decision tree.The regression tree is minimized by the square error, and the classification tree is selected by the Gini index minimization criterion to select the binary tree. Classification and regression trees are binary trees that can be used for classification and regression problems, as proposed by Breiman et al. [28]. The output of the classification tree is the category of the sample, and the output of the regression tree is a real number. The CART regression tree assumes that the tree is a binary tree, by constantly splitting features.

The objective function generated by the CART regression tree is shown in equation (5).

$$\sum_{x_i \in R_m} (y_i - f(x_i))^2. \quad (5)$$

In order to solve the optimal segmentation feature j and the optimal segmentation point s, the variable x^j is selected as the segmentation variable, and its value s is the segmentation point, then two regions R_1 and R_2 are obtained, such as the formula (6).

$$R_1(j,s) = \left\{ x \middle| x^j \le s \right\}, R_2(j,s) = \left\{ x \middle| x^j > s \right\}, \quad (6)$$

When j and s are fixed, the representative values c_1 and c_2 of the two regions are found to minimize the square difference between the respective intervals as in equation (7), and the objective function is converted into the formula (8).

$$\begin{aligned} &\min_{j,s}[\min_{c_1} \sum_{x_i \in R_1(j,s)} (y_i - c_1)^2 \\ &+ \min_{c_2} \sum_{x_i \in R_2(j,s)} (y_i - c_2)^2] \end{aligned}, \quad (7)$$

$$\begin{aligned} \widehat{c}_1 &= ave\left(y_i \middle| x_i \in R_1(j,s) \right), \\ \widehat{c}_2 &= ave\left(y_i \middle| x_i \in R_2(j,s) \right) \end{aligned}. \quad (8)$$

2.3.2 XGBoost

XGBoost is an improved algorithm based on gradient lifting decision tree, which can effectively build enhanced trees and run in parallel. The value of optimizing the objective function is at the core of XGBoost. The model corresponding to XGBoost contains multiple CART trees, so the XGBoost model can be expressed as a formula (9)

$$\hat{y} = \sum_{t=1}^{T} f_t(x_i), f_t \in F, \tag{9}$$

where T is the number of trees, F is all possible CART trees, and f_t is a specific CART tree.

The objective function of the XGBoost model can be expressed as formula (10).

$$\mathrm{Obj}^{(t)} = \sum_{i=1}^{n} l\left((y_i, \hat{y}^{(t-1)}) + f_t(x_i)\right) + \Omega(f_t) + C. \tag{10}$$

Among them, $\hat{y}^{(t-1)}$ means to retain the model prediction of the previous *t-1* round, f_t is a new function, and C is a constant term.

The target function is subjected to Taylor's second-order expansion, and two variables are defined for the purpose of the original objective function for convenience calculation. As shown in equation (11),

$$\begin{aligned} g_i &= \partial_{\hat{y}^{(t-1)}} l(y_i, \hat{y}^{(t-1)}) \\ h_i &= \partial^2_{\hat{y}^{(t-1)}} l(y_i, \hat{y}^{(t-1)}) \end{aligned}. \tag{11}$$

It can be seen that the objective function at this time can be changed to the form of formula (12).

$$\mathrm{Obj}^{(t)} \approx \begin{aligned} &\sum_{i=1}^{n}[l(y_i, \hat{y}^{(t-1)}) + g_i f_t(x_i) \\ &+ \frac{1}{2} h_i f_t^2(x_i)] + \Omega(f_t) + C \end{aligned}. \tag{12}$$

When the model is trained, the objective function can be expressed by formula (13).

$$\mathrm{Obj}^{(t)} = \sum_{j=1}^{t}[(\sum_{i\in I_j} g_i) w_j + \frac{1}{2}(\sum_{i\in I_j} h_i + \lambda) w_j^2] + \gamma T, \tag{13}$$

Define formula (14).

$$G_j = \sum_{i\in I_j} g_i, H_j = \sum_{i\in I_j} h_j. \tag{14}$$

Bring formula (14) into formula (13) to get formula (15).

$$\begin{aligned} \mathrm{Obj}^{(t)} &= \sum_{j=1}^{t}[G_j w_j + \frac{1}{2}(H_j + \lambda) w_j^2] + \gamma T \\ &= -\frac{1}{2}\sum_{j=1}^{T} \frac{G_j^2}{H_j + \lambda} + \gamma T \end{aligned}. \tag{15}$$

Equation (15), also known as the scoring function, is a measure of the quality of the tree structure. The smaller the value, the better the structure. We use the scoring function to select the best segmentation point to build a CART tree. Since the scoring function is a measure of the quality of the tree structure, the scoring function can be used to select the best segmentation point. First, all the cut points of the sample feature are determined, and each determined cut point is divided, and the standard of the good or bad is shown in formula (16).

$$\begin{aligned} Gain = \frac{1}{2}\Bigg[&\frac{G_L^2}{H_L + \lambda} + \frac{G_R^2}{H_R + \lambda} \\ &- \frac{(G_L + G_R)^2}{H_L + H_R + \lambda} \Bigg] - \gamma \end{aligned}, \tag{16}$$

Gain represents the difference between the single node $obj^{(t)}$ and the tree $obj^{(t)}$ of the two nodes after the segmentation. By traversing the segmentation points of all features, finding the largest segmentation point is the best segmentation point. According to this method, the node is continuously segmented to obtain the CART tree. When XGBoost predicts the nonlinear behavior of data sets, CART trees can explain the nonlinear relationship in the model and the interdependence between variables, and have good performance in nonlinear data prediction. Because the error part of DWT decomposition has nonlinear characteristics, this paper uses XGBoost model to predict the error part of stock data set. The predicted colleague used the grid search algorithm to optimize the parameters and construct the GSXGB model.

Grid search is an exhaustive search method for specifying parameter values. The optimal learning algorithm is obtained by optimizing the parameters of the estimation function through cross-validation. The possible values of each parameter are arranged and combined, and all the combined results are listed to generate a "grid". Each combined parameter is then used for XGBoost training and performance is evaluated using cross-validation. After the fitting function has tried all the parameter combinations, it returns a suitable classifier and automatically adjusts to the optimal parameter combination. To solve the uncertainty of random selection of parameter values. Firstly, according to the idea of the grid search algorithm, the parameter combination interval to be selected is first set. Based on the Xgboost algorithm, in the process of parameter optimization, combined with the idea of the grid search algorithm, the model is continuously trained, and each function is evaluated by the evaluation function. The classification results obtained by the combination of parameters are evaluated, and finally the optimal parameter combination is obtained. Finally, the optimal parameter combination is substituted into the Xgboost algorithm to construct GS-XGBoost, referred to as GSXGB.

The specific steps for building GSXGB are as follows:

(1) Obtaining a stock error partial data set of wavelet decomposition;

(2) Constructing the XGBoost prediction model;

(3) Apply the grid search algorithm to the XGBoost model, construct the GS-XGBoost prediction model, and train the GS-XGBoost model training using the error training set;

(4) Then test the GS-XGBoost prediction model using the test set;

(5) Compare the difference between the real result and the predicted result.

2.4 The hybrid methodology

The behavior of stock prices can not easily be captured. Hybrid models can simulate stock price behavior patterns to improve overall predictive performance. Therefore, a hybrid strategy that has stock prices forecasting modeling abilities is a good alternative for forecasting stock prices. Both the ARIMA and the XGBoost models have different capabilities to capture data characteristics in stock prices, so the hybrid model proposed in this study is composed of the ARIMA component and the XGBoost component. First, the stock data set (Y_t) is divided into approximate(L_t) and error(N_t) parts by discrete wavelet transform with basis function db4, as shown in formula (17), where t is time.

$$Y_t = L_t + N_t , \tag{17}$$

Then, for the separated data set L_t, the ARIMA (p, d, q) model is used to predict the training, and the approximation partial prediction data set (L_T) is obtained. The determination of p, d, q in the model is determined according to the ACF map and the PACF map, respectively. The data set N_t is predicted by the XGBoost model, and the error partial prediction data set

Table I. *The specific construction process of the model.*

Step 1: Import the historical data and perform missing value deletion processing.

    data = tushare.get_k_data(Stock code, start=start, end=end)
    data= data.replace(to_replace='?', value=np.nan).dropna()

Step 2: The discrete wavelet transform with the basis function db4 is used to denoise and decompose the stock historical data (Y_t) according to formula (17), and the opening price of the data set is divided into a approximation part (L_t) and a error part (N_t).

    L_t, N_t = pywt.wavedec(Y_t, 'db4', mode='sym', level=1)

Step 3: The ARIMA (p, d, q) model is used to train and predict the approximation part after decomposition, and the model p, d, q are determined according to the ACF map and the PACF map.

    params=ARIMA(L_t, order=(1, 1, 0)).fit()
    L_T= ARIMA(L_t, order=(1, 1, 0)).predict(params=params, start=1, end=len(L_t))

Step 4: Using the XGB method in the sklearn package, the XGBoost algorithm is implemented to train the decomposed error data sets. The XGBoost model uses the grid search algorithm to optimize the parameters and construct the GSXGB stock price prediction model.

    param_dist = {'max_depth': [x for x in range(1, 20, 1)], 'n_estimators': [x for x in range(1, 200, 1)], 'min_child_weight': [x for x in
                 range(1, 50, 1)], }
    N_T = GridSearchCV(XGBRegressor(max_depth, n_estimators, min_child_weight), param_grid, cv=5).fit().predict(N_t)

Step 5: The approximation part (L_T) predicted by the ARIMA (p, d, q) model and the error part (N_T)predicted by the GSXGB model are reconstructed according to the formula (18). The reconstructed result is (Y_T) the final prediction result.

    Y_T= pywt.waverec([L_T, N_T], 'db4')

Step 6: The prediction performance of the DWT-ARIMA-GSXGB model is judged by comparing the difference between the real value and the predicted value.

(N_T) is obtained. Since the parameters in the XGBoost algorithm affect the performance of the XGBoost algorithm, In order to solve the problem that the optimal parameters of XGBoost algorithm are difficult to find, this paper uses grid search algorithm (GS) to optimize the parameters, in which the CV in the GS is set to 5 based on human experience. Finally, the wavelet reconstruction is used to recombine L_T and N_T to obtain the final prediction result Y_T, as shown in formula (18).

$$Y_T = L_T + N_T. \tag{18}$$

The specific construction process of the model is shown in Table 1, and the flow chart is shown in figure 1.

III. Forecasting of Stock Prices

Ten stocks were used in this study to test the performance of the proposed model.Collect the daily opening price of each stock for 2015-2018 as the inventory data set.The last 100 data of each stock opening price is used as the verification data set, and the remaining data sets are used as training data sets.In this study, in order to prevent the estimation of the sample prediction error and the impact of the previous cumulative error on the prediction, only one step ahead prediction is considered. The specific information of the 10 stock data sets is shown in Table 2.

Five indices, RMSE (mean absolute percent error), MAE (mean absolute error), R^2 (determination coefficient), AUC, and Accuracy are used as a measure of prediction accuracy. The indices as follows

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y^{(i)} - Y^{(i)})^2}\ , \tag{19}$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|y^{(i)} - Y^{(i)}\right|, \tag{20}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - Y_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}\ , \tag{21}$$

$$AUC = \frac{\sum_{ins_i \in pos} rank_i - \frac{M \times (M+1)}{2}}{M \times N}\ , \tag{22}$$

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}\ . \tag{23}$$

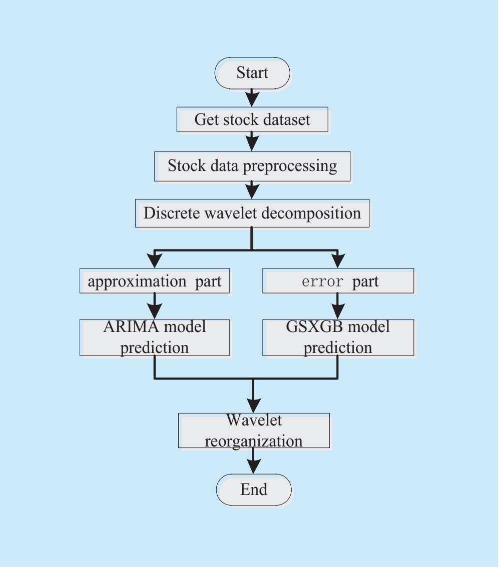


Fig. 1. *DWT-ARIMA-GSXGB model flow chart.*

Table II. *Data source information.*

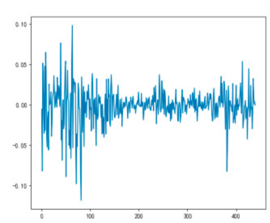
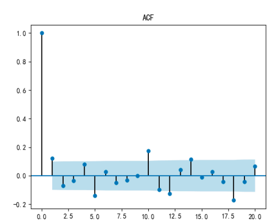
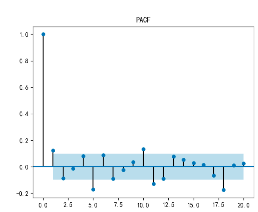
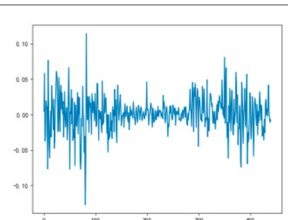
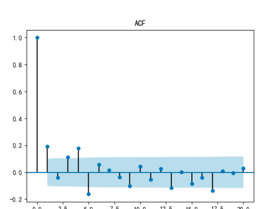
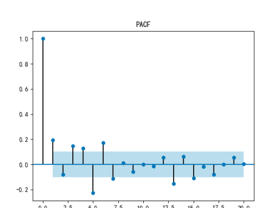
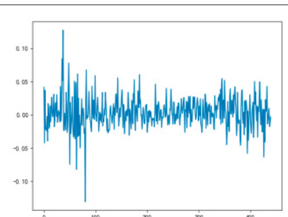
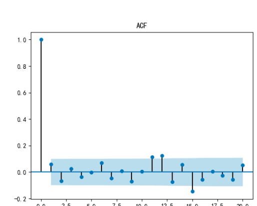
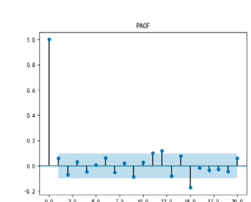
Stock name	China National Petroleum Corporation(CNPC)	PING AN INSURANCE (GROUP) COMPANY OF CHINA, LTD (PING AN)	Kweichow moutai Co., Ltd. (KMCL)	SAIC Motor Corporation Limited (SAIC Motor)	Foshan Haitian Flavouring and Food Co., Ltd. (HT-Food)	China State Construction Engineering Corporation(CSCEC)	S.F. Holding Co., Ltd.(S.F)	Fuyao Group	CASTECHINC.	CITIC SECURITIES CO., LTD. (CITIC)
Stock code	601857	601318	600519	600104	603288	601688	002352	600600	002222	600030
Stock date	1/5/2015-12/28/2018	1/5/2015-12/28/2018	1/5/2015-12/28/2018	1/5/2015-12/28/2018	1/5/2015-12/28/2018	1/5/2015-12/28/2018	1/5/2015-12/28/2018	1/5/2015-12/28/2018	1/5/2015-12/28/2018	1/5/2015-12/28/2018
Opening price	975	975	975	955	975	975	930	975	975	971

Among them, in the formula (19-21), y is a true value, and Y is a predicted value. In formula (22), $rank_i$ represents the serial number of the i-th stock data, and M and N are the number of positive samples (stock price rises) and negative samples (share price declines), respectively. In formula (23), TP and FP are the number of correct and wrong stocks predicted, and TN and FN are the number of correct and wrong stocks.

In this study, the stock historical data set is first divided into approximation and error parts by discrete wavelet transform with basis function db4. Then the ARIMA model is used respectively, and the improved XGBoost model predicts the approximation and error parts after the split. Finally, the wavelet reconstruction is used to combine the predicted results, and the combined prediction result is the final predicted value.The ARIMA model has three phases: model identification, parameter estimation, and diagnostic checking. Before fitting the ARIMA(p, d, q) model, you must specify the p, d, q of the model. The ACF map and the PACF graph contribute to the decision process of p and q in the ARIMA model.

The Data Plot part in Table 3 represents the data of the first-order difference of the approximate parts of the 10 sets of stock opening prices. It can be seen from the data part of the figure that the approximate part of the 10 stocks has basically stabilized, trends include increase/decrease trends, occasional large dip angles and stable correlation coefficients, and mixed oscillation periods, with most data sets showing oscillation trends that appear to be close to white noise. The ACF Plot and PACF Plot represent the autocorrelation and partial correlation graphs of the first 20 data, respectively. According to the ACF / PACF chart of the opening price of 10 sets of stock data, it can be judged that (p, d, q) = (1, 1, 0), (0, 1, 1), (2, 1, 0), (2, 1, 1), (3, 1, 0), (3, 1, 1), (3, 1, 2), (4, 1, 0), (4, 1, 1), (4, 1, 2) seems to be suitable for predicting all the data in this article. In order to better detect the performance of the hybrid model, the choice of the ARIMA optimal model is not considered here, but the val-

Table III. *Notable trends and the ACF/PACF of 1-level-differenced data.*

Stock Data	Data Plot	ACF Plot	PACF Plot
CNPC			
PING AN			
KMCL			

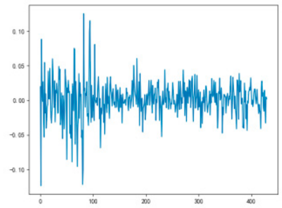
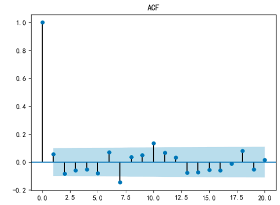
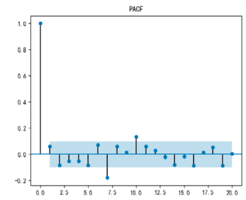
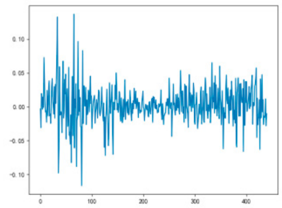
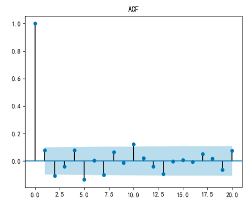
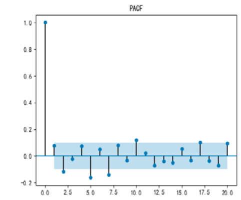
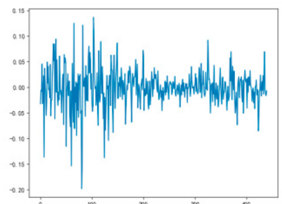
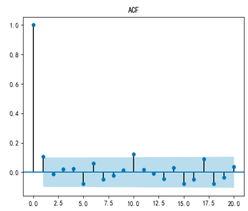
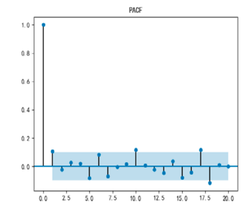
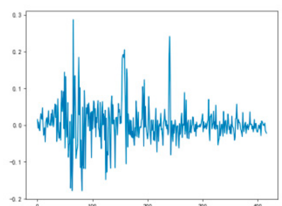
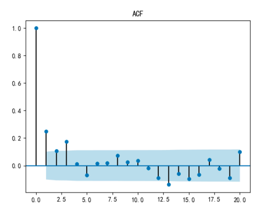
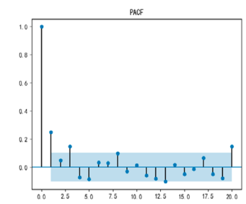
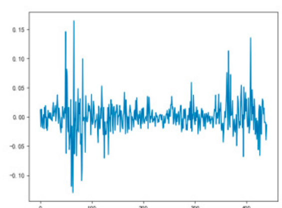
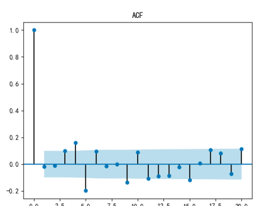
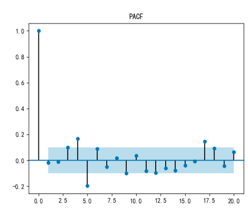
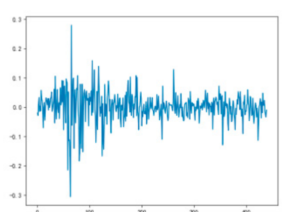
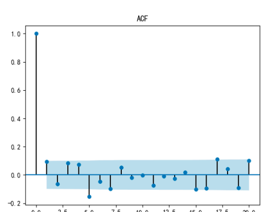
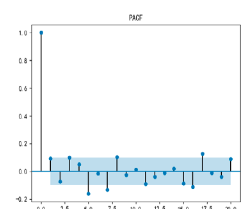
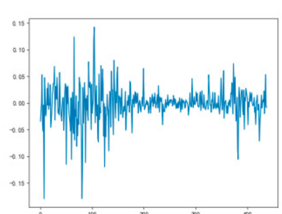
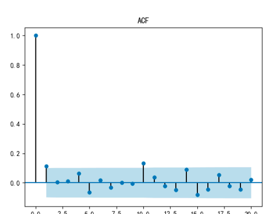
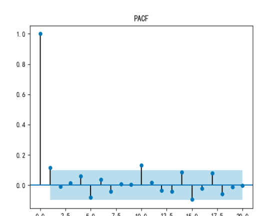
SAIC Motor			
HT-Food			
CSCEC			
S.F			
Fuyao Group			
CASTECHINC.			
CITIC			

Table IV. *Comparison of predictive indicators of DWT-ARIMA(0, 1, 1)-GSXGB model.*

	RMSE	MAE	P²	AUC	Accuracy
ARIMA(0, 1, 1) model					
1. CNPC	0.67501	0.78710	-2.0283	0.54582	54.55%
2. PING AN	471.001	21.2465	-54.547	0.48506	48.48%
3. KMCL	134707.	365.869	-45.662	0.46481	46.46%
4. SAIC Motor	108.213	10.3056	-29.141	0.53518	53.54%
5. HT-Food	1027.71	31.2866	-71.497	0.43428	43.43%
6. CSCEC	6.39467	2.12402	-3.6265	0.56591	56.57%
7. S.F	190.090	13.5408	-21.072	0.46481	46.46%
8. Fuyao Group	128.236	10.5442	-18.771	0.54551	54.55%
9. CASTECHINC	5.39963	2.01520	-5.7946	0.52512	52.53%
10. CITIC	1.63360	0.92211	-1.9490	0.57591	57.58%
XGBoost model					
1. CNPC	0.00539	0.05300	0.97581	0.73261	73.74%
2. PING AN	0.54991	0.57517	0.93514	0.72258	71.72%
3. KMCL	81.3232	7.08542	0.97182	0.64814	63.64%
4. SAIC Motor	0.13253	0.29011	0.96308	0.75490	75.76%
5. HT-Food	2.27602	1.14520	0.83944	0.56408	56.57%
6. CSCEC	0.05220	0.18098	0.96222	0.75714	75.76%
7. S.F	0.32150	0.41549	0.96266	0.56362	55.56%
8. Fuyao Group	0.22360	0.37932	0.96552	0.65714	65.66%
9. CASTECHINC	0.02889	0.13483	0.96363	0.71384	71.72%
10. CITIC	0.03054	0.13427	0.94486	0.72775	72.73%
GSXGB model					
1. CNPC	0.00388	0.04475	0.98255	0.75695	75.76%
2. PING AN	0.58152	0.59318	0.93141	0.72606	72.73%
3. KMCL	78.2017	6.62409	0.97256	0.70000	69.70%
4. SAIC Motor	0.13021	0.28482	0.96373	0.73772	73.74%
5. HT-Food	2.39683	1.19438	0.83092	0.54448	54.55%
6. CSCEC	0.04667	0.16319	0.96622	0.77816	77.78%
7. S.F	0.28832	0.40951	0.96652	0.61067	60.61%
8. Fuyao Group	0.19286	0.34883	0.97026	0.67653	67.68%
9. CASTECHINC	0.02893	0.13524	0.96358	0.72365	72.73%
10. CITIC	0.02854	0.12601	0.94847	0.73714	73.74%
DWT+ARIMA(0, 1, 1)+XGBoost model					
1. CNPC	0.00237	0.03740	0.98932	0.85720	85.86%
2. PING AN	0.35650	0.47561	0.95795	0.82835	82.83%
3. KMCL	36.3691	4.89102	0.98740	0.77777	77.78%
4. SAIC Motor	0.05886	0.19827	0.98360	0.80053	79.80%
5. HT-Food	0.48033	0.56755	0.96611	0.80795	80.81%
6. CSCEC	0.02662	0.13392	0.98073	0.77734	77.78%
7. S.F	0.14146	0.26035	0.98357	0.83694	83.84%
8. Fuyao Group	0.23138	0.40342	0.96432	0.73734	73.74%
9. CASTECHINC	0.02660	0.13404	0.96652	0.82843	82.83%
10. CITIC	0.02483	0.12650	0.95516	0.80775	80.81%
DWT + ARIMA(0, 1, 1) + GSXGB model					
1. CNPC	0.00174	0.03115	0.99218	0.87745	87.88%
2. PING AN	0.22161	0.37932	0.97386	0.87745	87.88%
3. KMCL	20.2947	3.66994	0.99296	0.79999	79.80%
4. SAIC Motor	0.03635	0.15905	0.98987	0.81014	80.81%
5. HT-Food	0.23330	0.41443	0.98354	0.83836	83.84%
6. CSCEC	0.01616	0.10369	0.98830	0.84836	84.85%
7. S.F	0.09425	0.21617	0.98905	0.89668	89.90%
8. Fuyao Group	0.11304	0.28878	0.98257	0.79775	79.80%
9. CASTECHINC	0.01710	0.10620	0.97848	0.87928	87.88%
10. CITIC	0.01416	0.09422	0.97443	0.81816	81.82%

Table V. *Comparison of predictive indicators of DWT-ARIMA(1, 1, 0)-GSXGB model.*

	RMSE	MAE	P^2	AUC	Accuracy
ARIMA(1, 1, 0) model					
1. CNPC	0.02936	0.13106	0.86824	0.64402	64.65%
2. PING AN	1.32716	0.89993	0.84348	0.65671	65.66%
3. KMCL	141.964	9.47687	0.95082	0.72222	71.72%
4. SAIC Motor	0.19049	0.34145	0.94694	0.69721	69.70%
5. HT-Food	1.44728	0.93487	0.89790	0.69693	69.70%
6. CSCEC	0.07837	0.20890	0.94329	0.68714	68.69%
7. S.F	0.14896	0.29985	0.98270	0.73874	73.74%
8. Fuyao Group	0.86317	0.69817	0.86691	0.72714	72.73%
9. CASTECHINC	0.07393	0.20649	0.90697	0.70649	70.71%
10. CITIC	0.07055	0.20318	0.87263	0.66632	66.67%
XGBoost model					
1. CNPC	0.00539	0.05300	0.97581	0.73261	73.74%
2. PING AN	0.54991	0.57517	0.93514	0.72258	71.72%
3. KMCL	81.3232	7.08542	0.97182	0.64814	63.64%
4. SAIC Motor	0.13253	0.29011	0.96308	0.75490	75.76%
5. HT-Food	2.27602	1.14520	0.83944	0.56408	56.57%
6. CSCEC	0.05220	0.18098	0.96222	0.75714	75.76%
7. S.F	0.32150	0.41549	0.96266	0.56362	55.56%
8. Fuyao Group	0.22360	0.37932	0.96552	0.65714	65.66%
9. CASTECHINC	0.02889	0.13483	0.96363	0.71384	71.72%
10. CITIC	0.03054	0.13427	0.94486	0.72775	72.73%
GSXGB model					
1. CNPC	0.00424	0.04464	0.98094	0.73874	73.74%
2. PING AN	0.58371	0.59601	0.93115	0.72606	72.73%
3. KMCL	79.2032	6.68972	0.97256	0.69074	68.69%
4. SAIC Motor	0.12607	0.28218	0.96488	0.73874	73.74%
5. HT-Food	2.48793	1.22223	0.82449	0.57551	57.58%
6. CSCEC	0.04312	0.15927	0.96880	0.79836	79.80%
7. S.F	0.28475	0.41577	0.96693	0.66837	66.67%
8. Fuyao Group	0.20067	0.35596	0.96905	0.67632	67.68%
9. CASTECHINC	0.03154	0.14312	0.96030	0.66789	66.67%
10. CITIC	0.02727	0.12423	0.95075	0.75693	75.76%
DWT + ARIMA(1, 1, 0) + XGBoost model					
1. CNPC	0.00240	0.03733	0.98919	0.84656	84.85%
2. PING AN	0.33199	0.45930	0.96084	0.82835	82.83%
3. KMCL	36.3691	4.89102	0.98740	0.77777	77.78%
4. SAIC Motor	0.05828	0.19787	0.98376	0.77925	77.78%
5. HT-Food	0.48167	0.56790	0.96602	0.80795	80.81%
6. CSCEC	0.03395	0.14335	0.98266	0.77734	77.78%
7. S.F	0.13170	0.25053	0.98470	0.83694	83.84%
8. Fuyao Group	0.25357	0.41413	0.96090	0.73734	73.74%
9. CASTECHINC	0.02633	0.13310	0.96686	0.83823	83.84%
10. CITIC	0.02482	0.12644	0.95518	0.80775	80.81%
DWT + ARIMA(1, 1, 0) + GSXGB model					
1. CNPC	0.00183	0.03186	0.99177	0.87847	87.88%
2. PING AN	0.32868	0.45127	0.96123	0.84758	84.85%
3. KMCL	17.6599	3.42838	0.99388	0.81851	81.82%
4. SAIC Motor	0.03579	0.15802	0.99003	0.81014	80.81%
5. HT-Food	0.09147	0.24848	0.96602	0.92897	92.93%
6. CSCEC	0.02666	0.13397	0.99354	0.83836	83.84%
7. S.F	0.06124	0.18284	0.99288	0.86579	86.87%
8. Fuyao Group	0.04759	0.17606	0.99266	0.87836	87.88%
9. CASTECHINC	0.01682	0.10477	0.97882	0.87928	87.88%
10. CITIC	0.01418	0.09421	0.97439	0.81816	81.82%

Table VI. *Comparison of predictive indicators of DWT-ARIMA(2, 1, 1)-GSXGB model.*

	RMSE	MAE	P^2	AUC	Accuracy
ARIMA(2, 1, 1) model					
1. CNPC	0.93132	0.92350	-3.1782	0.54582	54.55%
2. PING AN	635.427	24.6705	-73.939	0.48506	48.48%
3. KMCL	186842.	430.860	-63.722	0.46481	46.46%
4. SAIC Motor	154.924	12.3286	-42.152	0.53518	53.54%
5. HT-Food	1048.60	31.6009	-72.971	0.43428	43.43%
6. CSCEC	2.27300	1.20419	-0.64451	0.54551	54.55%
7. S.F	237.885	15.1482	-26.622	0.46481	46.46%
8. Fuyao Group	170.689	12.1630	-25.317	0.54551	54.55%
9. CASTECHINC	7.31028	2.34663	-8.1988	0.52512	52.53%
10. CITIC	1.68569	0.99995	-1.8971	0.59545	59.55%
XGBoost model					
1. CNPC	0.00539	0.05300	0.97581	0.73261	73.74%
2. PING AN	0.54991	0.57517	0.93514	0.72258	71.72%
3. KMCL	81.3232	7.08542	0.97182	0.64814	63.64%
4. SAIC Motor	0.13253	0.29011	0.96308	0.75490	75.76%
5. HT-Food	2.27602	1.14520	0.83944	0.56408	56.57%
6. CSCEC	0.05220	0.18098	0.96222	0.75714	75.76%
7. S.F	0.32150	0.41549	0.96266	0.56362	55.56%
8. Fuyao Group	0.22360	0.37932	0.96552	0.65714	65.66%
9. CASTECHINC	0.02889	0.13483	0.96363	0.71384	71.72%
10. CITIC	0.03054	0.13427	0.94486	0.72775	72.73%
GSXGB model					
1. CNPC	0.00412	0.04660	0.98148	0.76759	76.77%
2. PING AN	0.54595	0.58225	0.93141	0.70785	70.71%
3. KMCL	81.6389	7.09883	0.97172	0.64444	63.64%
4. SAIC Motor	0.12769	0.28507	0.96443	0.72810	72.73%
5. HT-Food	2.41562	1.20023	0.82959	0.55510	55.56%
6. CSCEC	0.04659	0.16267	0.96628	0.77795	77.78%
7. S.F	0.33958	0.45438	0.96056	0.61067	60.61%
8. Fuyao Group	0.25533	0.39315	0.96063	0.68673	68.69%
9. CASTECHINC	0.03196	0.14804	0.95977	0.67769	67.68%
10. CITIC	0.02854	0.12601	0.94847	0.73714	73.74%
DWT + ARIMA(2, 1, 1) + XGBoost model					
1. CNPC	0.00246	0.03829	0.98892	0.85720	85.86%
2. PING AN	0.38822	0.48371	0.95421	0.83797	83.84%
3. KMCL	36.6624	4.86288	0.98730	0.78888	78.79%
4. SAIC Motor	0.06293	0.20289	0.98247	0.78989	78.79%
5. HT-Food	0.56006	0.62002	0.96049	0.80816	80.81%
6. CSCEC	0.02680	0.13398	0.98060	0.83836	83.84%
7. S.F	0.13884	0.25705	0.98387	0.86783	86.87%
8. Fuyao Group	0.24130	0.40903	0.96279	0.73734	73.74%
9. CASTECHINC	0.02633	0.13459	0.96686	0.82843	82.83%
10. CITIC	0.02404	0.12178	0.95866	0.76388	76.40%
DWT + ARIMA(2, 1, 1) + GSXGB model					
1. CNPC	0.00183	0.03165	0.99177	0.87745	87.88%
2. PING AN	0.25428	0.39838	0.97386	0.87745	87.88%
3. KMCL	20.5723	3.66786	0.99287	0.79814	79.80%
4. SAIC Motor	0.04056	0.16687	0.98870	0.79950	79.80%
5. HT-Food	0.31242	0.47413	0.97796	0.83836	83.84%
6. CSCEC	0.01634	0.10402	0.98817	0.84836	84.85%
7. S.F	0.09404	0.21539	0.98908	0.89668	89.90%
8. Fuyao Group	0.11392	0.28894	0.98243	0.79775	79.80%
9. CASTECHINC	0.01690	0.10632	0.97873	0.86948	86.87%
10. CITIC	0.01424	0.09191	0.97551	0.83156	83.15%

Table VII. *Comparison of predictive indicators of DWT-ARIMA(3, 1, 0)-GSXGB model.*

	RMSE	MAE	P^2	AUC	Accuracy
ARIMA(3, 1, 0) model					
1. CNPC	0.03094	0.13428	0.86115	0.65364	65.66%
2. PING AN	3.99522	1.62830	0.52882	0.56403	56.57%
3. KMCL	384.751	16.0459	0.86672	0.70185	69.70%
4. SAIC Motor	0.78712	0.71621	0.78075	0.57467	57.58%
5. HT-Food	3.77732	1.48301	0.73353	0.43428	63.64%
6. CSCEC	0.39802	0.48888	0.71202	0.56571	56.57%
7. S.F	0.53290	0.58146	0.93812	0.62581	62.63%
8. Fuyao Group	2.77757	1.30040	0.57174	0.64653	64.65%
9. CASTECHINC	0.36459	0.47406	0.54120	0.56617	56.57%
10. CITIC	0.28792	0.40646	0.48023	0.56591	56.57%
XGBoost model					
1. CNPC	0.00539	0.05300	0.97581	0.73261	73.74%
2. PING AN	0.54991	0.57517	0.93514	0.72258	71.72%
3. KMCL	81.3232	7.08542	0.97182	0.64814	63.64%
4. SAIC Motor	0.13253	0.29011	0.96308	0.75490	75.76%
5. HT-Food	2.27602	1.14520	0.83944	0.56408	56.57%
6. CSCEC	0.05220	0.18098	0.96222	0.75714	75.76%
7. S.F	0.32150	0.41549	0.96266	0.56362	55.56%
8. Fuyao Group	0.22360	0.37932	0.96552	0.65714	65.66%
9. CASTECHINC	0.02889	0.13483	0.96363	0.71384	71.72%
10. CITIC	0.03054	0.13427	0.94486	0.72775	72.73%
GSXGB model					
1. CNPC	0.00395	0.04506	0.98227	0.75695	75.76%
2. PING AN	0.54794	0.57639	0.93537	0.72708	72.73%
3. KMCL	78.1873	6.85354	0.97291	0.64444	63.64%
4. SAIC Motor	0.12625	0.28177	0.96483	0.73874	73.74%
5. HT-Food	2.39310	1.19128	0.83118	0.57510	57.58%
6. CSCEC	0.04835	0.16927	0.96501	0.77816	77.78%
7. S.F	0.29923	0.41952	0.96525	0.62990	62.63%
8. Fuyao Group	0.23090	0.37282	0.96439	0.70673	70.71%
9. CASTECHINC	0.02589	0.12936	0.96741	0.74693	74.75%
10. CITIC	0.02859	0.12542	0.94837	0.77734	77.78%
DWT + ARIMA(3, 1, 0) + XGBoost model					
1. CNPC	0.00252	0.03761	0.98868	0.84656	84.85%
2. PING AN	0.39389	0.48846	0.95354	0.81669	81.82%
3. KMCL	36.4617	4.88869	0.98736	0.78888	78.79%
4. SAIC Motor	0.06011	0.19773	0.98325	0.77925	77.78%
5. HT-Food	0.48241	0.57075	0.96596	0.81795	81.82%
6. CSCEC	0.02718	0.13362	0.98033	0.83836	83.84%
7. S.F	0.14030	0.26006	0.98370	0.85822	85.86%
8. Fuyao Group	0.30944	0.46130	0.95228	0.71734	71.72%
9. CASTECHINC	0.02645	0.13466	0.96670	0.81862	81.82%
10. CITIC	0.02487	0.12645	0.95509	0.80775	80.81%
DWT + ARIMA(3, 1, 0) + GSXGB model					
1. CNPC	0.00188	0.03207	0.99153	0.86783	86.87%
2. PING AN	0.26000	0.40326	0.96933	0.87745	87.88%
3. KMCL	20.3013	3.67230	0.99296	0.79999	79.80%
4. SAIC Motor	0.03717	0.16133	0.98964	0.81014	80.81%
5. HT-Food	0.23419	0.41648	0.98347	0.83816	83.84%
6. CSCEC	0.01671	0.10384	0.98790	0.85836	85.86%
7. S.F	0.09321	0.21630	0.98917	0.89668	89.90%
8. Fuyao Group	0.16920	0.34245	0.97391	0.76816	76.77%
9. CASTECHINC	0.01693	0.10634	0.97869	0.84987	84.85%
10. CITIC	0.01421	0.09410	0.97433	0.82816	82.83%

ue of p, q is selected according to experience. In this paper, ARIMA (0, 1, 1), ARIMA (1, 1, 0), ARIMA (2, 1, 1) and ARIMA (3, 1, 0) random walk model are used to open the stock historical data set respectively. The approximate part is predicted. When the XGBoost model predicts the error part, the parameter selection affects the prediction result of the model. In order to achieve automatic optimization of parameters, the grid search algorithm is used to optimize the parameters max_depth, n_estimators and min_child_weight respectively.

In Table 4-7, The XGBoost model is consistent with the XGBoost parameter in the DWT-ARIMA-XGBoost model, with default parameters. Table 4 compares the predictions of the different models. These results show that the hybrid model DWT-ARIMA-GSXGB is superior to the ARIMA model, the XGBoost model, the grid search optimized XGBoost model (GSXGB) and the DWT-ARIMA-XG-

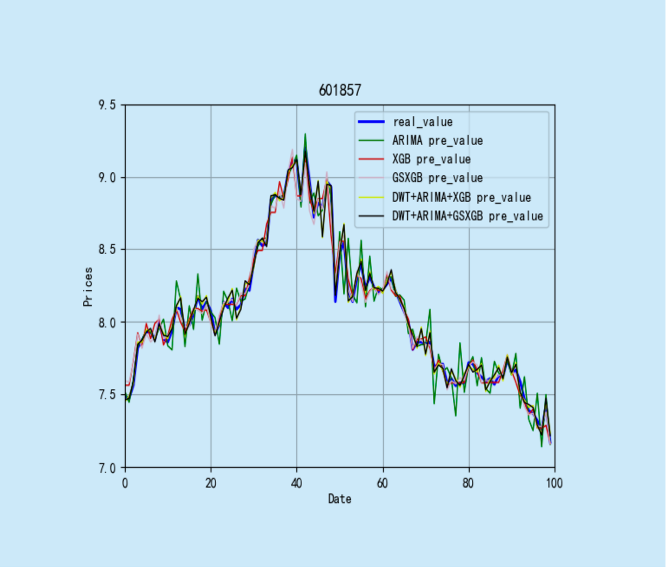


Fig. 2. *Stock prices of China National Petroleum Corporation.*

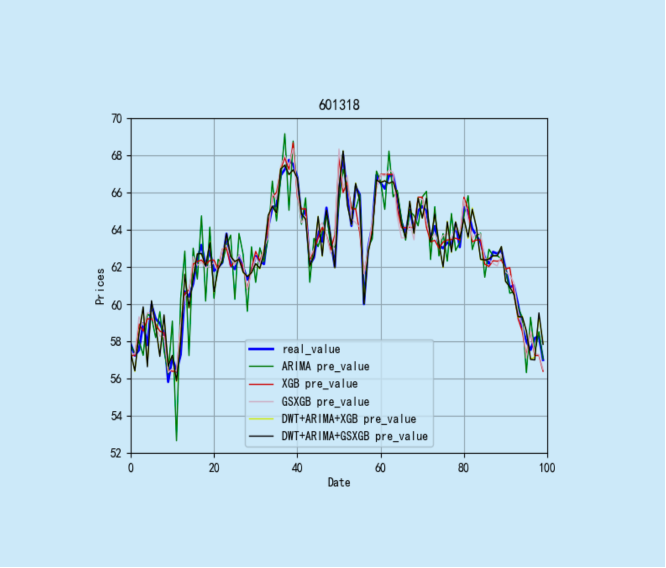


Fig. 3. *Stock prices of Ping An Insurance Company Of China.*

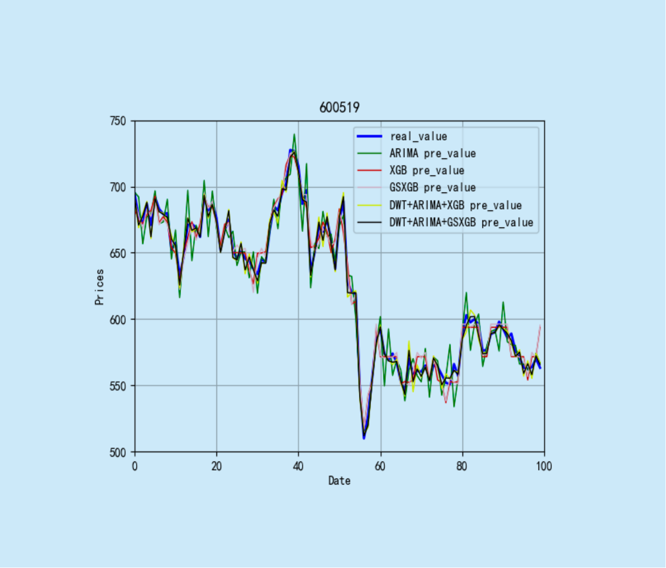


Fig. 4. *Stock prices of Kweichow Moutai Co., Ltd.*

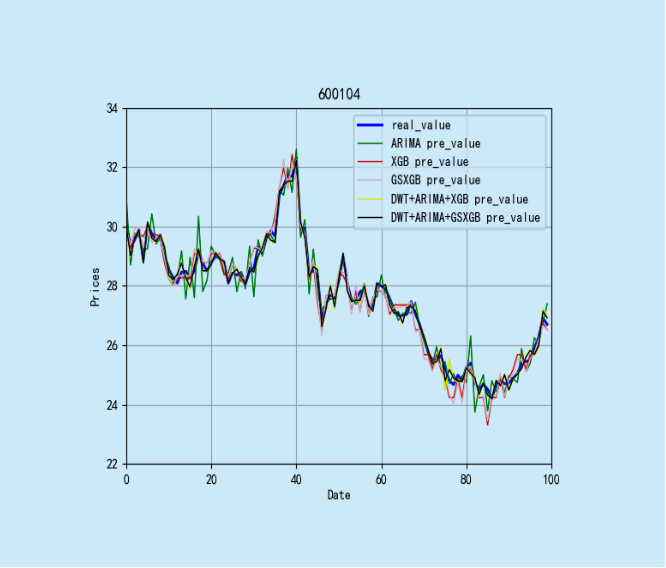


Fig. 5. *Stock prices of SAIC Motor Corporation Limited.*

Boost model in terms of five indices. In order to see the prediction effect of DWT-ARIMA-XGBoost more clearly, this paper only shows the comparison between the actual value and the predicted value of DWT-ARIMA(1, 1, 0)-GSXGB model. As shown in figure 2-11, it is clearer that the model proposed in this paper has a better fit in stock forecasting than several other comparative models. It further indicates that the DWT-ARIMA-GSXGB hybrid model proposed in this paper has higher fitness and predictive performance in stock price prediction than ARIMA model, XGBoost model, GSXGB model and DWT-ARIMA-XGBoost.

IV. Conclusion

The stock historical data set is a financial time series data set that can be decomposed into an approximate partial data set and an error partial data set. The ARIMA model dominates

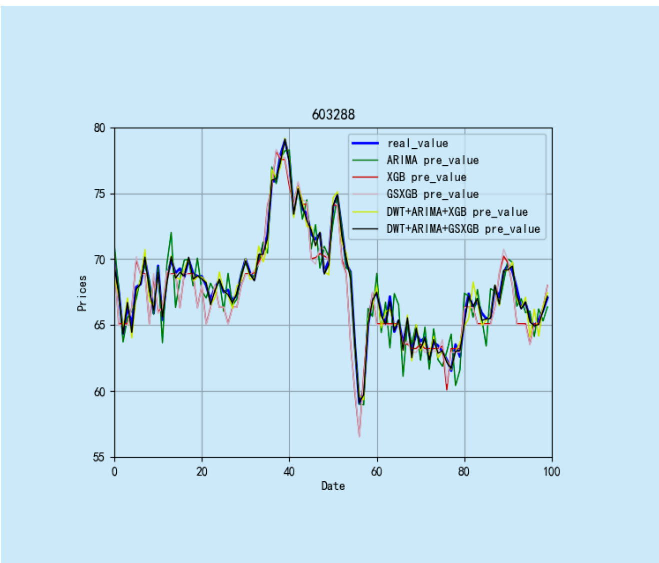


Fig. 6. *Stock prices of Foshan Haitian Flavouring and Food Co., Ltd.*

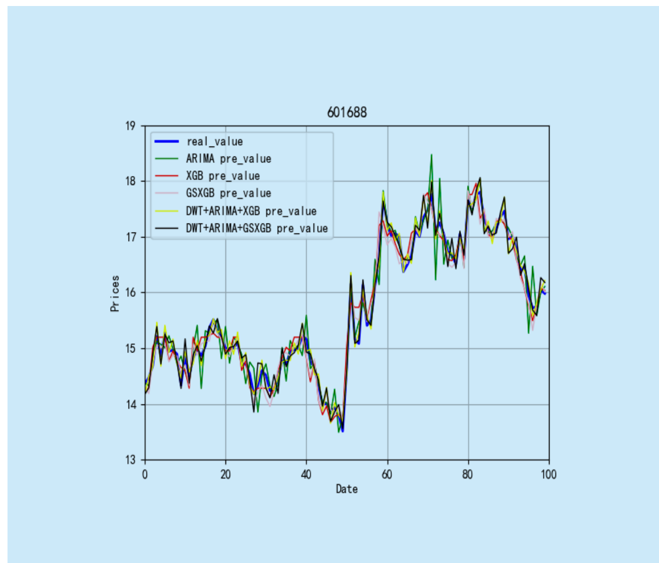


Fig. 7. *Stock prices of China State Construction Engineering Corporation.*

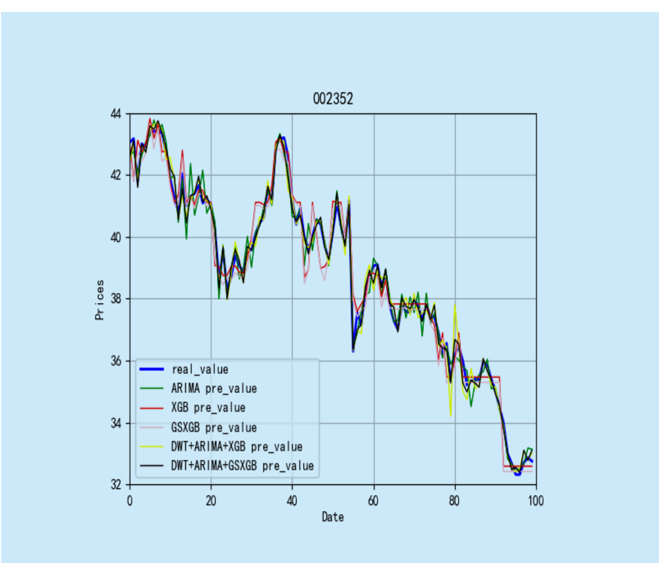


Fig. 8. *Stock prices of S. F. Holding Co., Ltd.*

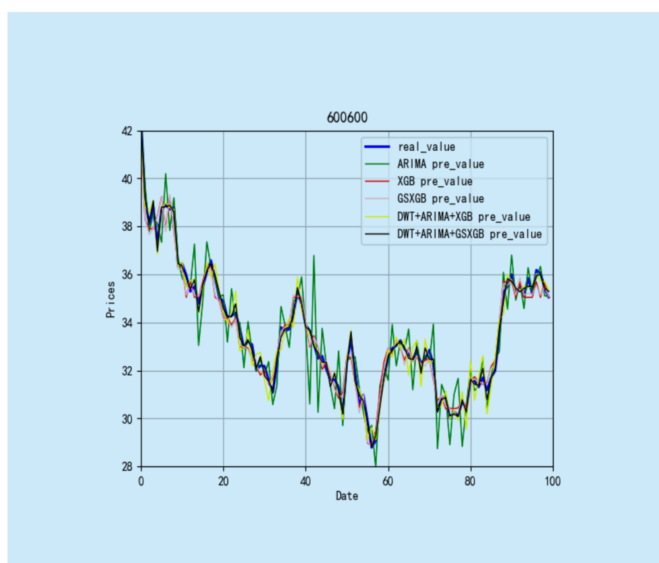


Fig. 9. *Stock prices of Fuyao Group.*

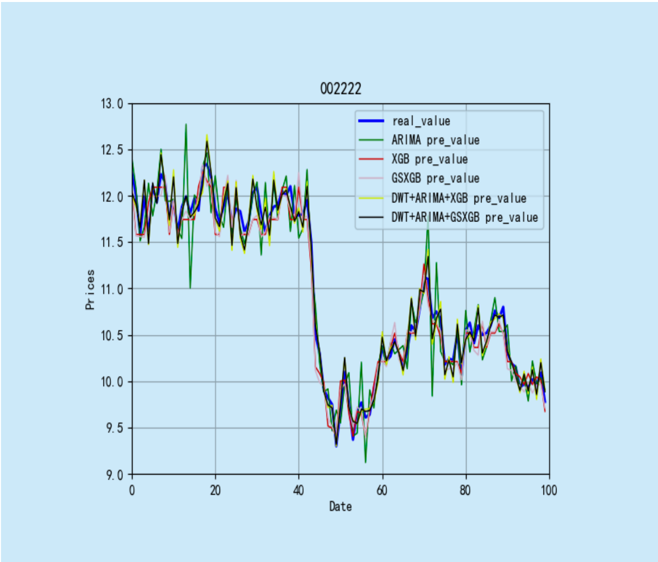


Fig. 10. *Stock prices of CASTECHINC.*

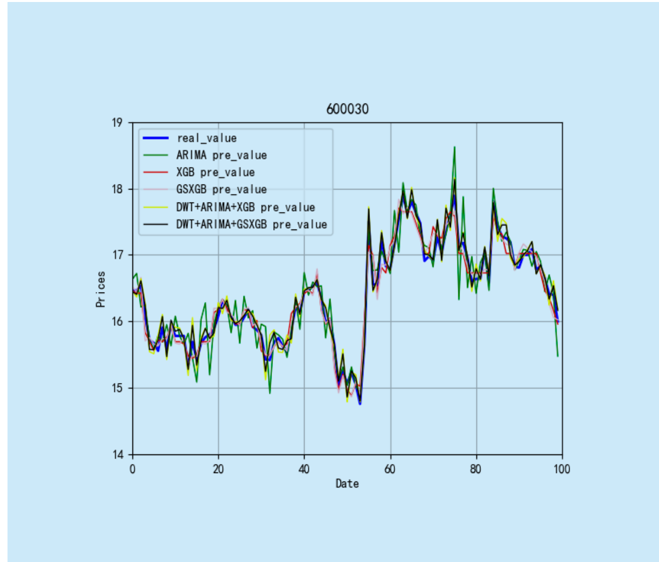


Fig. 11. *Stock prices of CITIC Securities Co., Ltd.*

many areas of time series prediction, which is basically a data-oriented approach that applies to the structure of the data itself. However, any error partial data set limits the predictive performance of the model. The CART tree in XGBoost can explain the nonlinear relationship between models and the interdependence between variables, and has good performance in nonlinear data prediction. Because the error part of DWT decomposition has nonlinear characteristics, this paper uses XGBoost model to predict the error part of stock data set. Therefore, this paper proposes a DWT-ARIMA-GSXGB hybrid model. Firstly, the discrete datalet transform is used to split the data set into approximation and error parts. Then the ARIMA model processes the approximation data and the improved xgboost model (GSXGB) handles error data. Finally, the prediction results are combined using wavelet reconstruction. According to the experimental comparison of 10 stock data sets, it is found that the errors of DWT-ARIMA-GSXGB model are less than the four prediction models of ARIMA, XGBoost, GSXGB and DWT-ARIMA-XGBoost. The simulation results show that the DWT-ARIMA-GSXGB stock price prediction model has good approximation ability and generalization ability, and can fit the stock index opening price well. And the proposed model is considered , . Theoretically and empirically, hybridization of two different models can reduce prediction errors. However, the diversity of model parameters makes that a simple combination of the two best individual models does not necessarily produce the best results. Therefore, the structural selection of the optimal parameters of the hybrid model has important research significance.

References

[1] Mcqueen G, Thorley S. "Are stock returns predictable? A test using markov chains." *Journal of Finance*, vol. 46, no. 1, 1991, pp. 239-263.

[2] Zhang G, Zhang X, Feng H. "Forecasting financial time series using a methodology based on autoregressive integrated moving average and Taylor expansion." *Expert Systems*, vol. 33, no. 5, 2016, pp. 501-516.

[3] Baldauf B, Santoni G J. "Stock price volatility: Some evidence from an ARCH model." *Journal of Futures Markets*, vol. 11, no. 2, 2010, pp. 191-200.

[4] Yamaguchi K. "Reexamination of stock price reaction to environmental performance: A GARCH application." *Ecological Economics*, vol. 68, no. 1, 2008, pp. 345-352.

[5] Pyo S, Lee J, Cha M, et al. "Predictability of machine learning techniques to forecast the trends of market index prices: Hypothesis testing for the Korean stock markets." *Plos One*, vol. 12, no. 11, 2017, pp. e0188107.

[6] Ming-Tao, Wu , and Y. Yong . "The research on stock price forecast model based on data min-

ing of BP neural networks." *Intelligent System Design and Engineering Applications (ISDEA), 2013 Third International Conference on* IEEE, 2013.

[7] Bao W, Yue J, Rao Y. "A deep learning framework for financial time series using stacked autoencoders and long-short term memory." vol. 12, no. 7, *Plos One*, 2017, pp. e0180944.

[8] Liu J N K, Hu Y. "Application of feature-weighted Support Vector regression using grey correlation degree to stock price forecasting." *Neural Computing & Applications*, vol. 22, no. 1, 2013, pp. 143-152.

[9] Chang T S. "A comparative study of artificial neural networks, and decision trees for digital game content stocks price prediction." *Expert Systems with Applications*, vol. 38, no. 12, 2011, pp. 14846-14851.

[10] Kim K J . "Financial time series forecasting using support vector machines." *Neurocomputing*, vol. 55, no. 1-2, 2003, pp. 307-319.

[11] Otengabayie E F, "Frimpong Magnus J. modelling and forecasting volatility of returns on the Ghana stock Exchange." *American Journal of Applied Sciences*, vol. 3, no. 10, 2006.

[12] Laboissiere L A, Fernandes R A S, "Lage G G. maximum and minimum stock price forecasting of Brazilian power distribution companies based on artificial neural networks." *Applied Soft Computing*, vol. 35, 2015, pp. 66-74.

[13] Adebiyi, Ayodele Ariyo , A. O. Adewumi , and C. K. Ayo . "Stock price prediction using the ARIMA model. "*2014 UKSim-AMSS 16th International Conference on Computer Modelling and Simulation* IEEE, 2014.

[14] Bates J M, Granger C W J. "The Combination of Forecasts." *Journal of the Operational Research Society*, vol. 20, no. 4,1969, pp. 451-468.

[15] Clemen R. "Combining forecasts: a review and annotated bibliography with discussion." *International Journal of Forecasting*, vol. 5, 1989, pp. 559–608.

[16] Menezes Lilian M, Bunn Derek W, "Taylor James W. Review of guidelines for the use of combined forecasts." *European Journal of Operational Research*, vol. 120, 2000, pp. 190–204.

[17] Lam KF, Mui HW, Yuen HK. "A note on minimizing absolute percentage error in combined forecasts." *Computer &Operations Research*, vol. 28, 2001, pp. 1141–7.

[18] Zhang G P. "Time series forecasting using a hybrid ARIMA and neural network model." *Neurocomputing*, vol. 50, no. 1, 2003, pp. 159-175.

[19] Zhang, Yan , et al. "A new wavelet-neural network-ARIMA shares index combination forecast model." *International Conference on Automatic Control & Artificial Intelligence* IET, 2013.

[20] Shi S, Liu W, Jin M. "Stock price forecasting based on a combined model of ARMA and BP neural network and Markov model." *International Journal of Information Processing & Management*, 2013, 4(3):215-221.

[21] Ye, Tian. "Stock forecasting method based on wavelet analysis and ARIMA-SVR model." *2017 3rd International Conference on Information Management*. 2017, pp. 102-106.

[22] Zhu H, Wei H, Jing Y, et al. "Soil organic carbon prediction based on scale-specific relationships with environmental factors by discrete wavelet transform." vol. 330, Geoderma, 2018, pp. 9-18.

[23] Chen T, Guestrin C. "XGBoost: A Scalable Tree Boosting System." *CoRR abs*, 2016, pp. 1603-02754.

[24] Almeida L B. "The fractional Fourier transform and time-frequency representations." *IEEE Transactions on Signal Processing*, vol. 42, no. 11, 1994, pp. 3084-3091.

[25] Sinha S, Routh P S, Anno P D, et al. "Spectral decomposition of seismic data with continuous-wavelet transform." *Geophysics*, vol. 70, no. 70, 2008.

[26] Pankratz, Alan . *Forecasting with Univariate Box-Jenkins Models: Concepts and Cases*. 2008.

[27] De'Ath G, Fabricius K E. "Classification and regression trees: a powerful yet simple technique for ecological data analysis." *Ecology*, vol. 81, no. 11, 2000, pp. 3178-3192.

[28] Grajski K A, Breiman L, Prisco G V D, et al. "Classification of EEG spatial patterns with a tree-structured methodology: CART." *IEEE Trans Biomed Eng*, vol. 33, no. 12, 1986, pp. 1076-1086.

Biographies

Yan Wang, is a professor at Lanzhou University of Technology and a master's tutor. Her current research interests include pattern recognition, image processing, and intelligent information processing.

Yuankai Guo, is currently pursuing the degree insoftware engineering with Lanzhou University of Technology, Lanzhou, China. His research interestsinclude time series analysis, machine learning, and data mining.