

A DFS Model for Forecasting Stock Price

Xiaolu Li^{1,a}, Hanghang Zhao^{2,b}, Kaiqiang Zheng^{3,c}, Shuaishuai Sun^{4,d}

¹Beijing University of Posts and Telecommunications, Beijing, P.R.China

²Beijing University of Posts and Telecommunications, Beijing, P.R.China

³Beijing University of Posts and Telecommunications, Beijing, P.R.China

⁴Beijing University of Posts and Telecommunications, Beijing, P.R.China

^axllill3932@163.com, ^bzhaohanghang@bupt.edu.cn, ^c1695130462@qq.com,

^dsun_2017@bupt.edu.cn

Keywords: DFS model; time series model; combination forecast model; wavelet analysis; Fourier transform; fitting analysis

Abstract. Currently, forecasting stock price is the hotter topic for achieving the smallest lost in investment. However, the previous stock price forecasting model practically cannot satisfy the requirement of accuracy.

To raise the forecast accuracy, a decomposition-forecast-synthesis (*DFS*) model is proposed by this paper, based on the analysis of the characteristics of the stock price time series, combined with the established single stock price prediction model, for instance, time series model, grey prediction model, neural network prediction model, etc.

DFS model decomposes a stock price time series into three components, including time trend component, quasi periodic component and random component. For each component, an adaptive prediction model is adopted to predict, afterwards, the synthesis of three component is to acquire stock price forecast sequence.

The wavelet analysis, combination forecast method, Fourier Transform, fitting analysis, and conventional time series models, for instance, the ARMA (Autoregressive-moving-average) method and exponent smoothness method are adopted in the *DFS* model.

For evaluation the forecasting performance of proposed model, the daily stock prices of SAIC Motor from December 2, 2013 to March 2, 2016 are used as experimental dataset and the Mean Square Error (MSE) and Mean Square Coefficient of Variation (MCV) as evaluation criterion.

1. Introduction

The stock market is an economic “barometer” and “alarm”. While for stock investors, the more accurate the future stock price forecasts, the lower risk and more secure are expected to obtain from the stock investment, the stock prediction exploration also plays an important role for a country’s economic advancement. Therefore, the study of stock intrinsic properties and prediction is of great theoretical significance and application prospect.

Because the stock information data is complex, computer-aided data processing is expected to apply to analyze and process the large amount of stock information data, instead only relying on experience and intuition of the relevant experts or practitioners for judgment. In this paper, on the basis of the existing stock prediction model, a *DFS* model is proposed via computer aided processing and data mining.

2. *DFS* Model

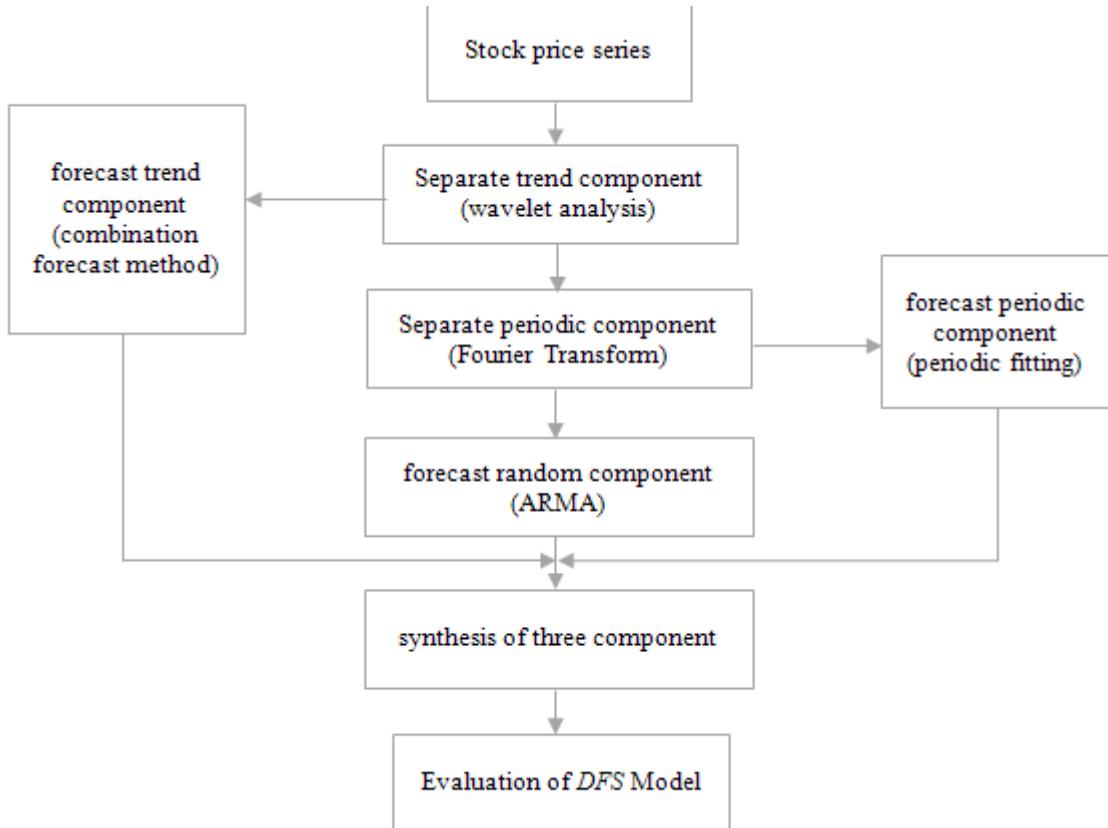
In terms of the characteristics of stock price series, it is a time series, therefore the methods of time series analysis are able to apply. After analysis of stock price time series, the characteristics of the time sequence of overall trend, certain period of time and small random fluctuation are discovered. Virtually, the internal factors affect the characteristics of the three aspects. The evolution of social economic for a considerable tim

Impact the overall trends, certain periodic reflects stock price changing with seasons, and small fluctuation may be due to short-term economic fluctuations and the psychological factors of investment.

To raise the forecast accuracy, a decomposition-forecast-synthesis (*DFS*) model is proposed by this paper, based on the analysis of the characteristics of the stock price time series, combined with the established single stock price prediction model, for instance, time series model, grey prediction model, neural network prediction model, etc.

DFS model decomposes a stock price time series into three components, including time trend component, quasi periodic component and random component. For each component, an adaptive prediction model is adopted to predict, afterwards, the synthesis of three component is to acquire stock price forecast sequence.

A. *DFS* Model Flow Chart.



B. *DFS* Model Algorithm. For easy understanding the proposed method, five steps are introduced step by step as follows:

Step 1: Separate trend component. Due to the growing trend of the signal often represents the low frequency components, the application of wavelet analysis is able to decompose the signal in multi-scale and then obtain the development trend from the low frequency coefficient. The primary reason is that, as the scale of wavelet decomposition increasing, the more high frequency components will be filter out, so that acquiring the development trend of the signal.

Wavelet analysis is applied to separate trend component of the stock price sequence. Above all, carry out five scale wavelet decomposition of sequence, then, reconstruct low frequency coefficient of wavelet decomposition, ultimately, select one scale in five as the composition of the trend of stock price series.

Step 2: forecast trend component. Got the trend of stock price series can carry on the forecast. Because the trend sequence is a smooth sequence removed high frequency information, abundant suitable prediction methods are able to forecast trend sequence, for instance, time series forecasting model, grey forecasting model, regression analysis prediction method, etc.

Single prediction methods can achieve the limited prediction precision, what's more, different methods suit different stages of different characteristics of the trend. To achieve higher prediction precision overall, apply the combination forecast model in this paper.

- **Combination Forecast Model.** Supposing m single forecasting methods, the prediction by i th single method is $f_i, i = 1, 2, \dots, m$.

If the combination prediction f meet $f = l_1 f_1 + l_2 f_2 + \dots + l_m f_m$, naming the combination forecast as linear combination forecast model, among l_1, l_2, \dots, l_m represent the weighting coefficient for each method, generally $\sum_{i=1}^m l_i = 1, l_i \geq 0, i = 1, 2, \dots, m$.

The core matter of combination forecast is to calculate the weighing coefficient, to effectively improve the prediction accuracy. The Prediction Error Sum of Squares Inverse Method is adopted.

- **Prediction Error Sum of Squares Inverse Method.** The greater prediction error sum of squares, the lower precision of the forecasting model, the smaller weighting coefficient in the combination forecast model.

Command:

$$l_i = E_{ii}^{-1} / \sum_{i=1}^m E_{ii}^{-1}, i = 1, 2, \dots, m \quad (1)$$

Absolutely, $\sum_{i=1}^m l_i = 1, l_i \geq 0, i = 1, 2, \dots, m$, among E_{ii} representing the sum of forecasting error square of i th single method.

$$E_{ii} = \sum_{t=1}^N e_{it}^2 = \sum_{t=1}^N (x_t - x_{it})^2. \quad (2)$$

x_{it} represents the predicted value of i th single method in t th time, x_t represents the actual value sequence $\{x_t, t = 1, 2, \dots, N\}$ of the same forecast object, N represent the length of time, $e_{it} = (x_t - x_{it})$ represents the forecasting error of i th single method in t th time.

- In this paper, the single forecasting methods adopt time series forecasting model, grey prediction model and polynomial fitting model, the weighted coefficient determined by Prediction Error Sum of Squares Inverse Method.

Step 3: Separate and forecast periodic component. Obtain the rest ingredient after removing the trend component from the stock price series. Applying Fourier Transform in rest ingredient to acquire the frequency practically energy concentrated, believed represented the roughly period of time.

Then fit the rest ingredient via cycle fitting to gain the quasi periodic fitting equation used to separate and forecast periodic component.

Step 4: forecast almost random component. Obtain the almost random sequence primary containing high frequency components after removing the periodic component from the rest ingredient. The autoregressive moving average (ARMA) model is applied to forecast the high frequency almost random sequence, one of the time series prediction model.

Step 5: synthesis of three component. Currently, the predicted value of trend component, periodic component and almost random component has been obtained. Then the synthetic process is to sum the three component predicted value to acquire the eventual predicted value of stock price series.

Step 6: Evaluation of DFS Model. The Mean Square Error (MSE) and Mean Square Coefficient of Variation (MCV) are adopted as evaluation criterion. The smaller MSE and MCV reflect the higher precision of forecasting model.

- **Mean Square Error (MSE).** The filtering theory told, if a forecast is sound, it is unbiased in statistical sense, meaning the expectation of forecast value is expected the true value:

$$MSE = \sqrt{\frac{\sum_{i=1}^n (\tilde{p}_i - E(\tilde{p}_i))^2}{n}} = \sqrt{\frac{\sum_{i=1}^n (\tilde{p}_i - p_i)^2}{n}} \quad (3)$$

- **Mean Square Coefficient of Variation (MCV).** To eliminate the dimension influence on the results of evaluation, adopting dimensionless mean square from the rate the MCV:

$$MCV = \frac{MSE}{\bar{P}} \quad (4)$$

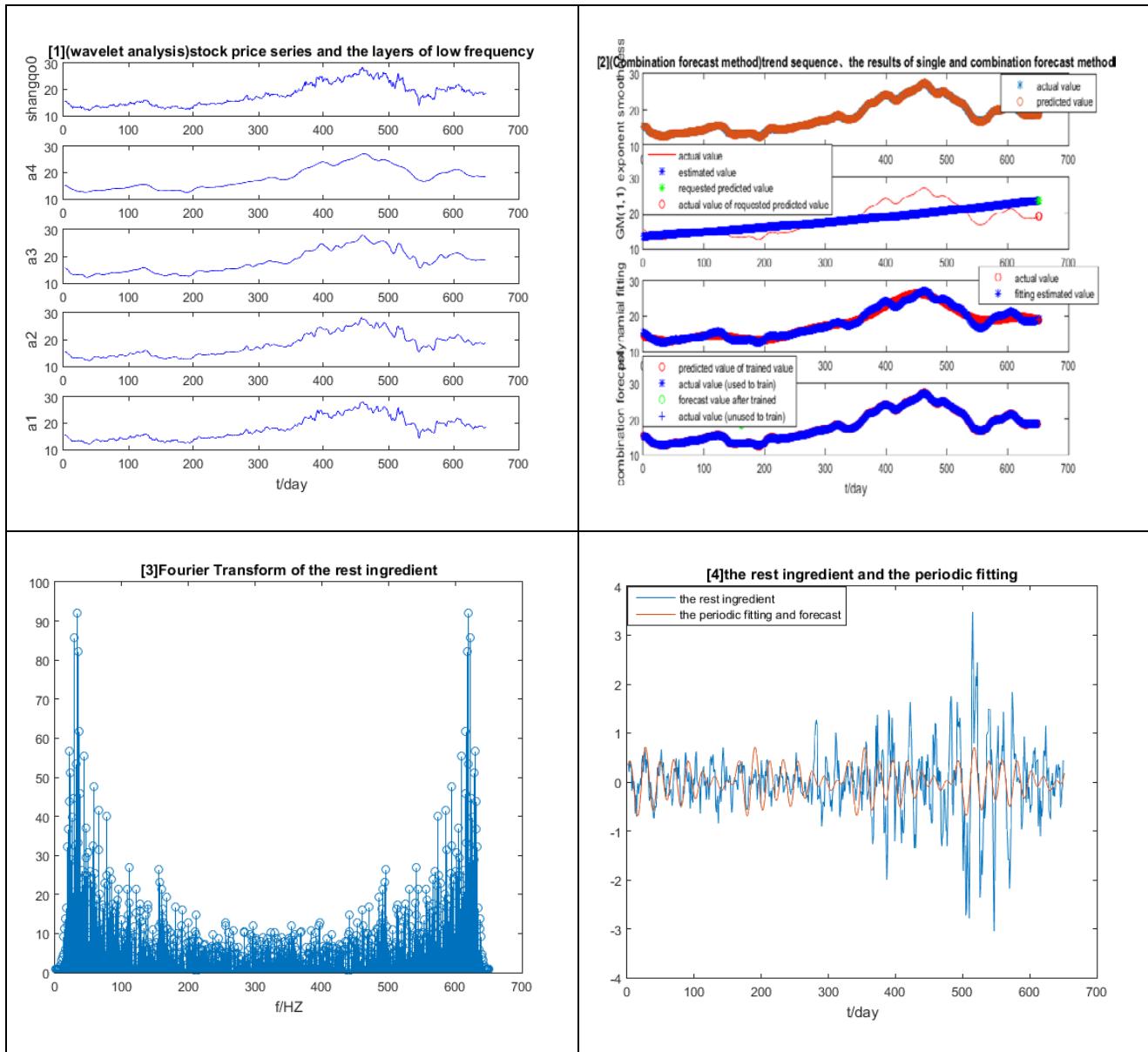
It is promoted by the standard deviation coefficient Cv ($Cv(x) = \frac{\sqrt{D(X)}}{E(X)}$), expressing the degree between the predicted and actual value. $\bar{P} > 0$, \bar{P} represented the expectation price of one stock, \tilde{p}_i and p_i respectively represent the predicted value and the actual value.

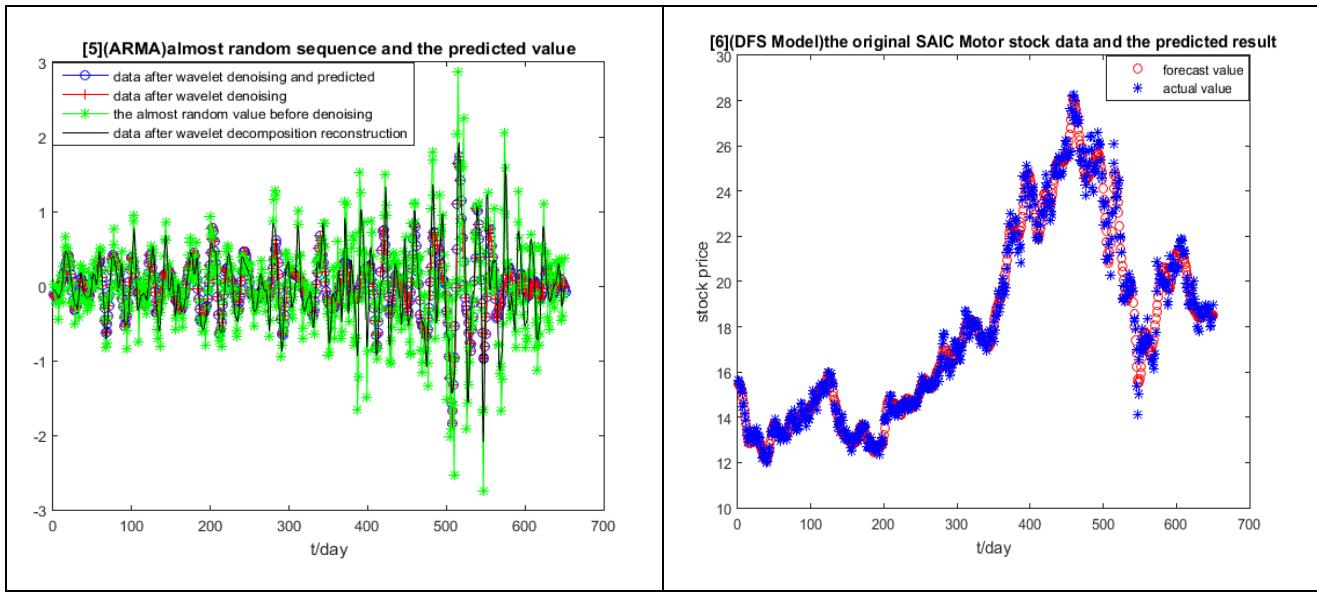
3. Experimental Results.

For evaluation the forecasting performance of proposed model, the daily stock prices of SAIC Motor from December 2, 2013 to March 2, 2016 are used as experimental dataset. Even more, a single forecast model is adopted to forecast the SAIC Motor stock, then comparing the two model. Here, the grey prediction model (GM (1, 1)) is adopted.

- **DFS Model**

Table 1 DFS Model





● GM(1,1)

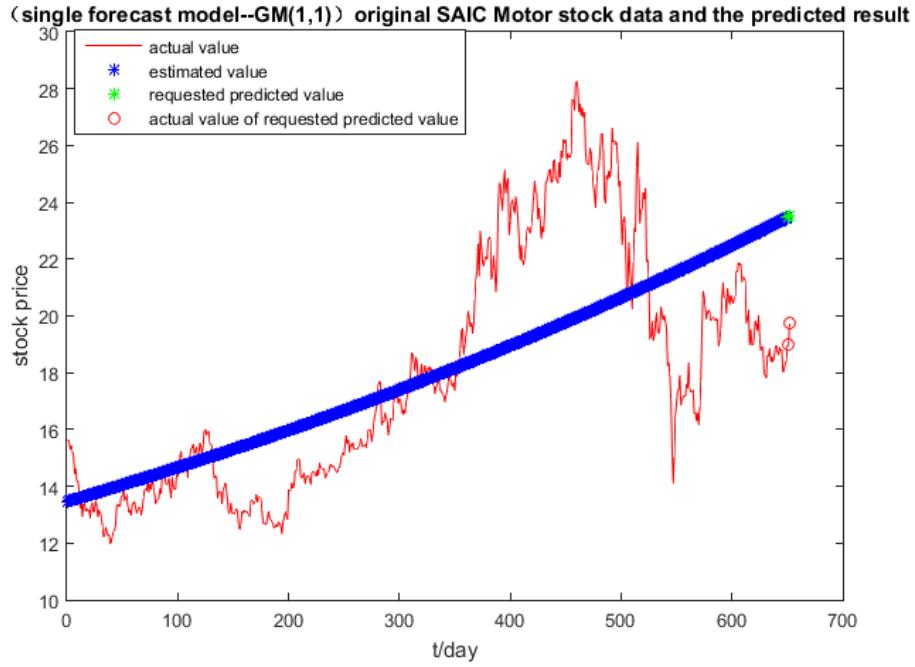


Fig 1 GM(1,1)

● Comparison of the two model

Table 2 Comparison of the two model

Evaluation Criterion	DFS	GM(1,1)
MSE	0.9816	6.0628
MCV	0.0547	0.3376

By contrast the evaluation criterion of *DFS* and GM(1,1), *DFS* Model achieved a higher precision satisfactory result.

4 Conclusion

The advantage of the *DFS* Model can be achieved from the experiment above compared to other single prediction model. It is true that the *DFS* Model addresses on the stock price series prediction

problem with various growing characteristics, for it is based on the analysis of the characteristics of stock price time sequence, combined with the existing prediction method, established a fresh stock price prediction model.

Acknowledgements

This work is supported by the Research Innovation Fund for College Students of Beijing University of Posts and Telecommunications (No.151).

References

- [1] G. E. P. Box, and G. M. Jenkins, "Time series analysis: Forecasting and control," Holden-Day, San Francisco, CA, USA, 1976.
- [2] L. Cao, and F. E. H. Tay, "Support vector machine with adaptive parameters in financial time series forecasting," IEEE Transactions on Neural Networks, vol. 14 (6), pp. 1506–1518, 2003.
- [3] G. Valeriy, and B. Supriya, "Support vector machine as an efficient framework for stock market volatility forecasting", Computational Management Science, vol. 3 (2), pp. 147–160, 2006.
- [4] C. Y. Yeh, C. W. Huang, and S. J. Lee, "A multiple-kernel support vector regression approach for stock market price forecasting", Expert Systems with Applications, vol. 38 (3), pp. 2177–2186, 2011.
- [5] Kazema, E. Sharifia, F. K. Hussainb, M. Saberic, and O. K. Hussaind, "Support vector regression with chaos-based firefly algorithm for stock market price forecasting," Applied soft computing, vol. 13, pp. 947-958, 2013.
- [6] F. Tsai, Y. C. Lin, D. C. Yen, and Y. M. Chen, "Predicting stock returns by classifier ensembles", Applied Soft Computing, vol. 11(2), pp. 2452-2459, 2011.
- [7] Ediger, and S. Akar, "ARIMA forecasting of primary energy demand by fuel in Turkey," Energy Policy, vol. 35, pp. 1701-1708, 2007.
- [8] J. H. Friedman, "Multivariate adaptive regression splines (with discussion)," Ann. Stat, vol. 19, pp. 1–141, 1991.
- [9] Huang, Alex Yihou,"Asymmetric dynamics of stock price continuation", "Journal of Banking and Finance", vol. 36, no. 6, pp. 1839-1855, 2012.
- [10] Johnstone, D.J, "Economic Interpretation of Probabilities Estimated by Maximum Likelihood or Score", "Management Science", vol.57, no.2, pp.308-314, 2011