Optimal Forecast Combination Based on Ensemble Empirical Mode Decomposition for Agricultural Commodity Futures Prices

Yongmei Fang[1,2,3], Bo Guan[3] , Shangjuan Wu[1] , and Saeed Heravi[3]

[1]College of Mathematics and Informatics, South China Agricultural University,
[2]College of Economics and Management, South China Normal University,
[3] Cardiff Business School, University of Cardiff, UK

Abstract:
Improving the prediction accuracy of agricultural product futures prices is important for the investors, agricultural producers and policy makers. This is to evade the risks and enable the government departments to formulate appropriate agricultural regulations and policies. This study employs Ensemble Empirical Mode Decomposition (EEMD) technique to decompose six different categories of agricultural futures prices. Subsequently three models, Support Vector Machine (SVM), Neural Network (NN) and ARIMA models are used to predict the decomposition components. The final hybrid model is then constructed by comparing the prediction performance of the decomposition components. The predicting performance of the combination model were then compared with the benchmark individual models, SVM, NN, and ARIMA. Our main interest in this study is on the short-term forecasting, and thus we only consider 1-day and 3-days forecast horizons. The results indicated that the prediction performance of EEMD combined model is better than that of individual models, especially for the 3-days forecasting horizon. The study also concluded that the machine learning methods outperform the statistical methods to forecast high-frequency volatile components. However, there is no obvious difference between individual models in predicting the low-frequency components.

Keywords: Forecast combination, Hybrid model; Future prices. Support Vector Machine

Corresponding author: Saeed Heravi; Heravis@cardiff.ac.uk

1. Introduction

The first standardized futures contract in China was made in |May 1993 for Wheat. 1993. There are currently around 20 categories of agricultural futures listed in China, with the trading volume of 978 million and turnover of 34.89 trillion yuan.

Chinese agricultural futures market has significant impact on the world futures market, with the market share of 58% in the trading volume in the global agricultural market in 2011. As reported in the United States Futures Association in 2014, half of the top 20 trading volume of agricultural futures and options products are from China. Among them, vegetable, soybean, sugar, natural rubber, and palm oil ranked in the top 5 products and soybean oil, eggs, cotton, yellow soybean and rapeseed oil listed the 7^{th}, 9^{th}, 10^{th}, 13^{th} and 18^{th}, respectively. The trading volume of the above ten categories amounted to more than 940 million, which is approximately 70% of the total trading volume of global agricultural futures and options. Therefore, the Chinese agricultural futures play an increasingly important role in the international market, and hence accurate forecast of prices is vital for producers and investors.

The available studies on forecasting futures prices have been focused on the crude oil (Wei, 2012; Kang & Yoon, 2013; Sévi, 2014; Barunik & Malinská, 2016; Wen, Gong & Cai, 2016; Fileccia & Sgarra, 2018; Liu, Yang & Zhang, 2018; Ma et. al, 2018; Yang, Gong & Zhang, 2018), precious metals (Wei, 2009; Yang & Dai, 2013, Li & Li, 2015; Luo & Ye, 2015; Lin & Gong, 2017; Bonato *et. al*, 2018; Fang & Xiao, 2018; Fang, Yu & Xiao, 2018), stock index (Hamid & Iqbal, 2004; Chu et. al, 2009; Yang & Liu, 2014) , and carbon (Byun & Cho, 2013). The influence of crude oil on the agricultural futures prices have also been examined and the risks associated with agricultural future prices are investigated in (Huang, Huang & Wang, 2013; Yang & Tian, 2014a, 2014b; Cartwright & Riabko, 2015; Tian & Yang, 2016; Tian, Yang & Chen, 2017; Yang et. al, 2017).

With the increasing share of China's agricultural futures in the international market, increasingly more scholars began to pay attention to Chinese agricultural futures (Li & Lu, 2011; Bohl, Siklos & Wellenreuther, 2018). Xiong *et al.* (2015) applied the VECM-MSVR technique to predict interval prices for the Chinese agricultural futures, and showed that the linear and non-linear information of the time series can be captured better by combination models. Teng & Zhou (2017) and Chu (2014) employed ARMA and ARMA (2,2)-Garch (1,1) models and Liang & Tai (2014) used the EGARCH-EWMA model to forecast soybean futures prices. Gao & Yu (2014) predicted cotton futures prices, using the EGARCH-EWMA and ARIMA models, and concluded that the performance of the EGARCH-EWMA model was better than that ARIMA. Teng & Zhou (2017), Chen & Huang (2010) compared econometric models with ARIMA and showed that ARIMA could achieve better results in the short-term forecasting.

With the developments of the theory and practice of artificial intelligence, these methods were also extensively employed in forecasting. For example, the wavelet method was applied to forecast the international crude oil prices and convolution neural network and back propagation neural network were employed to forecast the prices of zinc (Lin & Gong, 2017; Yousefi, Shahriar et al., 2005; Hamid & Iqbal, 2004). Yang & Dai (2013) optimized the SVM by the improved fish swarm algorithm, and predicted non-ferrous metal prices, showing improvement in the short term foresting. Zhang (2012) further found that the multi-variate least squares support vector machine outperforms the uni-variate method in predicting maize prices.

With the popularity and increasing use of combination models, ARIMA and EGARCH-EWMA model were applied by Gao & Yu (2014) to forecast the short-term prices of cotton. Xiong *et.al* (2015) implemented the vector error correction model to predict the linear feature of the futures prices, and multi-output support vector regression to fit the non-linear feature of cotton and corn prices in Zhengzhou Commodity Exchange in China. The research showed that the combination model outperforms the individual models. The combination model of ARIMA and LSSVM were implemented by Wang (2015) to predict carbon prices, again proving the advantage of using the combination methods. Yang & Liu (2014) used SYM8 wavelet to reduce the noise in the data, then BP neural network was trained and tested both on the de-noised and raw data. The result showed that the reducing noise could improve the forecast accuracy of stock index futures significantly.

In this paper, the Ensemble Empirical Mode Decomposition (EEMD) technique was used to decompose the data into linear and non-linear characteristics, then different prediction models were applied on the decomposed components, choosing the best model for each component. The results of the combination model are then compared with the individual models of SVM, NN and ARIMA, as the benchmark models. The research framework of this paper is organized as follows. In Section 2, we briefly describe the hybrid models for forecasting time series data. Section 3 describes the agricultural futures prices data used in the study. The results of combination model and individual models are analyzed and compared in Section 4, conclusions are drawn in the final section.

2. Methodology

2.1 Ensemble Empirical Mode Decomposition

In this section, we first briefly describe the original Empirical Mode Decomposition (EMD), which is proposed by Yeh et. al (2010). EMD is an adaptive method suitable for effectively capturing non-stationary and non-linear behavior in time series data. EMD decomposes the time series into n Intrinsic Mode Functions (IMF) with different frequency and amplitude, and a reminder as follows:

1. Determine the maximum (minimum) values of the original time series.
2. Apply a cubic interpolation and connect all the maximum (minimum) to generate

the upper(lower) envelope.

3. Obtain the local mean values of the two envelopes

$$m_1(t) = \left[x_{\max}(t) + x_{\min}(t)\right]/2 \tag{1}$$

4. Subtract the means obtained in (1) from the original time series data

$$h_1(t) = x(t) - m_1(t) \tag{2}$$

5. If $h_1(t)$ satisfies the IMF conditions, then repeat step 1 to step 4 until the remainder becomes a monotonic function and no more IMF can be extracted, in which the series is decomposed into n independent IMFs and a remainder,

$$x(t) = \sum_{i=1}^{n} h_i(t) + r(t)$$

An improved Ensemble Empirical Mode Decomposition (EEDM) is proposed by Wu & Huang (2009) to avoid aliasing produced by empirical mode decomposition, by adding noise to the data set. The process of ensemble empirical model decomposition is as following.

1. A white noise series confirming to normal distribution $\varepsilon_n(t)$ is added to the original time series, which generates a new time sequence as:

$$x_n(t) = x(t) + \varepsilon_n(t) \tag{3}$$

2. Decompose the time series data obtained in (3) into IMFs.
3. Repeat step 1 and step 2 m-times, with adding different white noise series.
4. As the final result, compute the averages of the corresponding IMFs obtained in the decomposition, step 2.

$$h_n(t) = \frac{1}{m}\sum_{i=1}^{m} h_{in}(t)$$

The advantage of EEMD is that the added noise cancel each other in the end results and the chance of mode mixing is significantly reduced. The final decomposition result is given as:

$$x(t) = \sum_{i=1}^{n} h_i(t) + r(t)$$

Where $h_i, i = 1, 2, ...n$ are the final IMFs and r is the remainder.

The intrinsic model functions and the remainder obtained by ensemble empirical model decomposition preserve the non-stationary and non-linear features of the original time series data while avoid the modal aliasing.

2.2 Support Vector Machine

Support Vector Machine (SVM) is a new machine learning method and has been widely used in many fields. SVM can deal with practical problems effectively, such as small sample problem, non-linear regression and high dimension pattern recognition Vapnik, (1998).By using the pre-selected kernel function, the input data is mapped into a high dimensional feature space. Then the optimal classification plane which maximizes the distance between the hyperplane is constructed from this high dimensional feature space. Support vector machine method can be used in the linear and non-linear forecasting, the fitting equation is given below:

$$f(x) = w^T x + b \tag{4}$$

Where w is the weight coefficients and b is the offset item. The weight parameters and the offset term can be obtained by introducing the Lagrangian function and find the optimized solution for the non-linear regression given below:

$$f(x) = \sum_{i=1}^{n} (\overline{a_i} - \overline{a_i^*}) k(x_i x) + \bar{b}$$

Where $\overline{a_i}$, $\overline{a_i^*}$, $\bar{b}$ are the optimum solution for the parameters and $k(x_i x)$ is the kernel function.

2.3 Neural Network

Neural network was first proposed by Rumelhant and Mcllelland in 1986 as a forward multilayer backward propagation network. Its structure includes input layer, hidden layer and output layer, and obtains the best fit to the data by adjusting the weights and thresholds of neural network nodes Rumelhart & McClleland (1986).The NN estimation procedure is briefly described below:

1. Compute the input signal s_k of the k hidden layer neuron by a weighted combination of all the inputs, i.e.

$$S_k = \sum_{k=1}^{m} w_{ks} x_s + d_k$$

Where w_{ks} indicates the k^{th} neuron weight of each neuron in input layer, and d_k is the k^{th} neuron's threshold.

2. Calculate the output value y_k from the hidden layer neuron node k as

$$y_k = f(S_k)$$

Where f is activation function of the hidden layer.

3. The error signal is transmitted back to each neuron through the network according to the original connection path, and the weight w between each neuron node and the threshold d of the node are modified continuously until the output result meets the expected result.

2.4 Autoregressive Integrated Moving Average Model

The autoregressive integrated moving average model ARIMA(p,d,q) is an extension of the autoregressive moving average model ARMA(p,q) proposed by Box & Jenkins (1970), where d is the differencing parameter and p and q are orders of the lags.

$$x_t = \varphi_0 + \varphi_1 x_{t-1} + \varphi_2 x_{t-2} + \cdots + \varphi_p x_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \cdots + \theta_q \varepsilon_{t-q}$$

Because most of the time series are non-stationary, it is necessary to transform the non-stationary time series data into the stationary time series by the d-order differencing and then apply ARMA model fitting procedure. The parameters p, q are chosen by Akaike information criterion (AIC) or Bayesian information criterion (BIC) in this paper.

3. The Data and EEMD Decomposition

The data used in this study are taken from Wind database. Six categories of agricultural futures are selected: vegetable meal, soybean meal, stalked rice, strong wheat, Zheng Cotton and early Indica rice. The data of futures prices are the daily closing prices and in all cases, our sample period starts from 27th November 2014 and ends in 11th October 2017. The number observations, presented in Figure 1, for each category is 700. The summary statistics of the data are given in Table 1.

Table 1: Descriptive Statistical of Future Prices

Category	Mean	Standard deviation	Minimum	Maximum	C.V.
Vegetable meal	2167.1	199.29	1529	2746	9.2%
Soybean meal	2783.1	224.73	2323	3486	8.8%
Stalked rice	3113.9	140.18	2762	3490	4.5%
Strong wheat	2697.9	72.915	2566	2992	2.6%
Early Indica rice	2575.0	206.89	1948	3096	8.0%
Zheng Cotton	13798.5	1716.64	10070	16880	12.4%

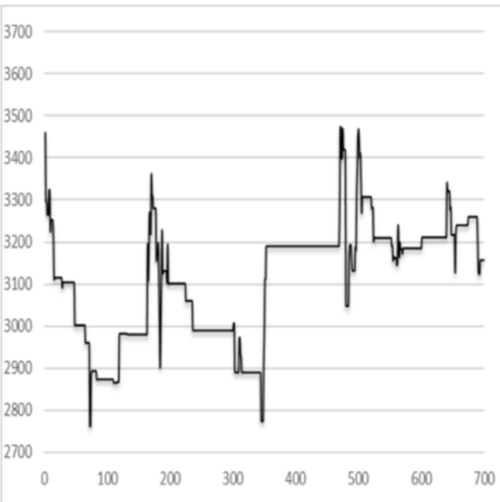
As can be seen from table 1, the standard deviation of the futures prices ranging from 72.91 for strong wheat to 1716.64 for Zheng Cotton. The highest coefficient of variation is also obtained for Cotton as 12.4%, which indicates much more volatility and fluctuation for Zheng Cotton than the other five products. The average price of cotton also was 4 to 6 times higher than the other five products in this period.

In this study, we utilize the logarithm of future prices, with 80% of the data (560 observations) used for modeling, the remaining 20% (140 observations) were used for testing. The EEMD method is used to decompose the transformed data. Figure 2 shows the decompositions of the six categories of future prices.

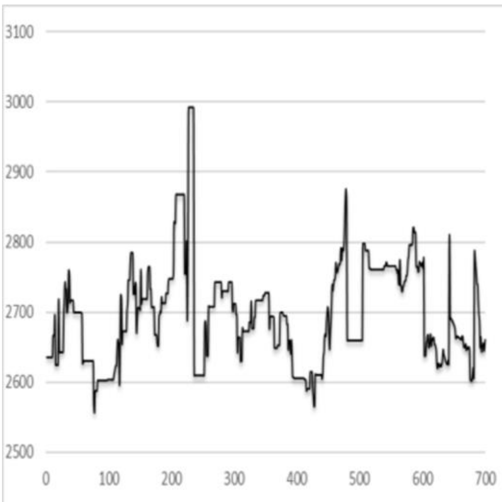
It can be seen from Figure 2 that the futures prices of six categories of agricultural commodities are decomposed into eight intrinsic model function components (from high to low frequency) and a remainder. The fluctuation period reflects the time length and the amplitude reflects the magnitude of the shock on the futures prices. The remainder displaying a monotonous increasing trend determines the long-term trend of future prices, which accords with the termination conditions of ensemble empirical mode decomposition.

a. vegetable meal

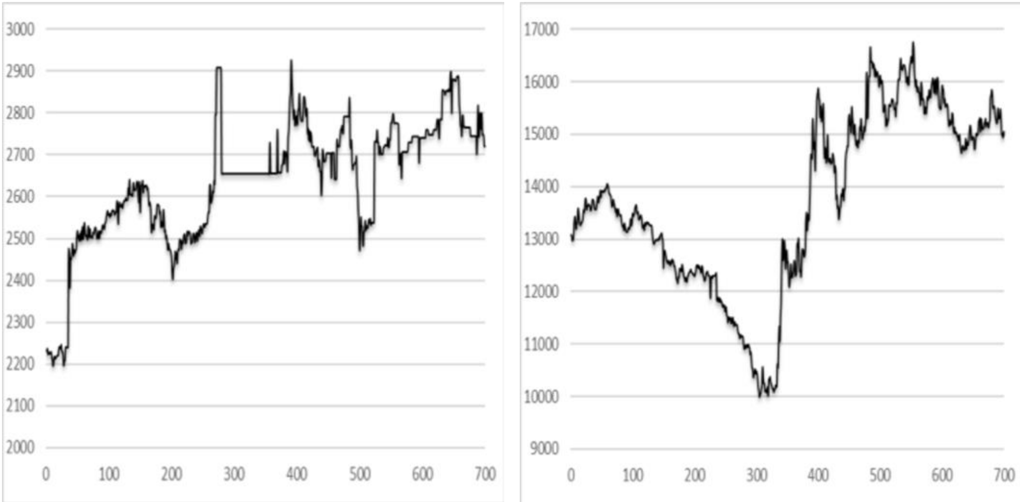
b. soybean meal



c. stalked rice



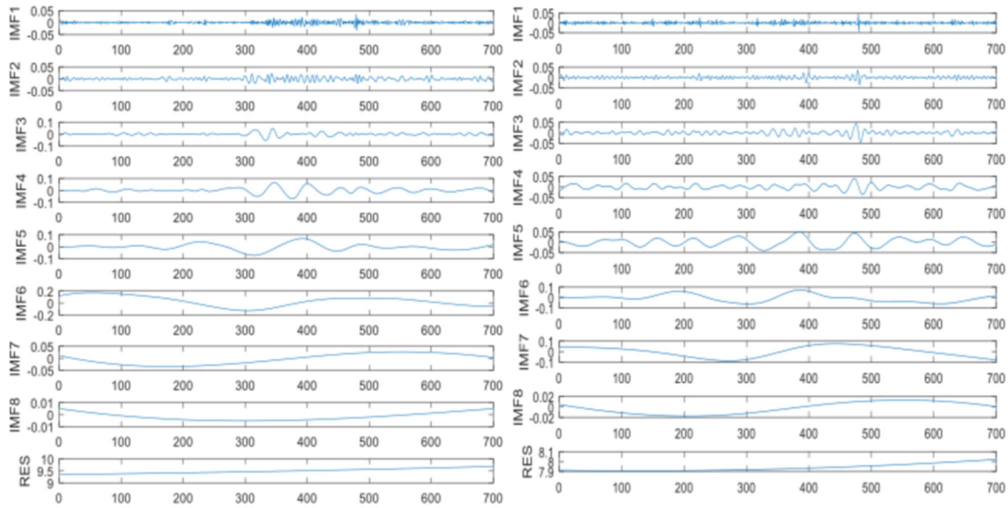
d. strong wheat



e. early Indica rice

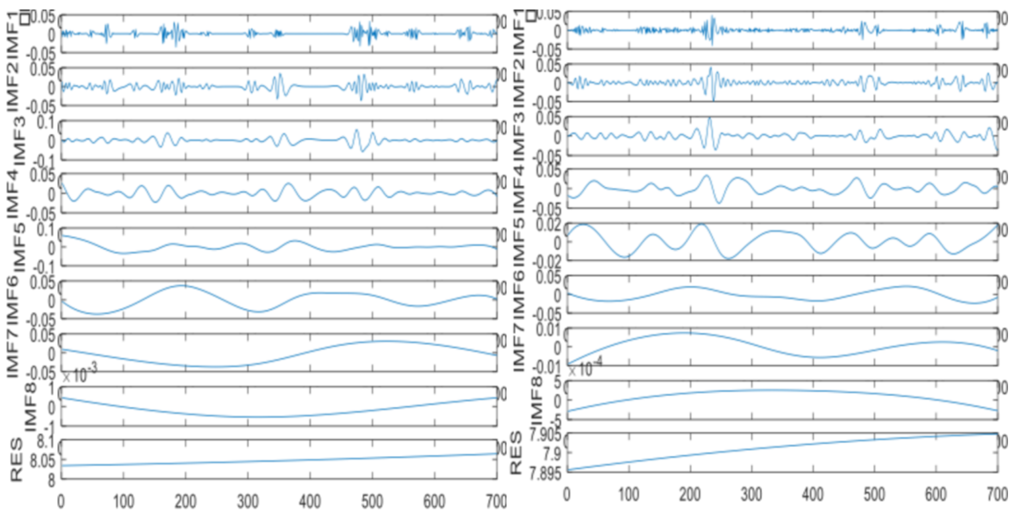
f. Zheng Cotton

Figure 1: Future Prices for the Six Commodities

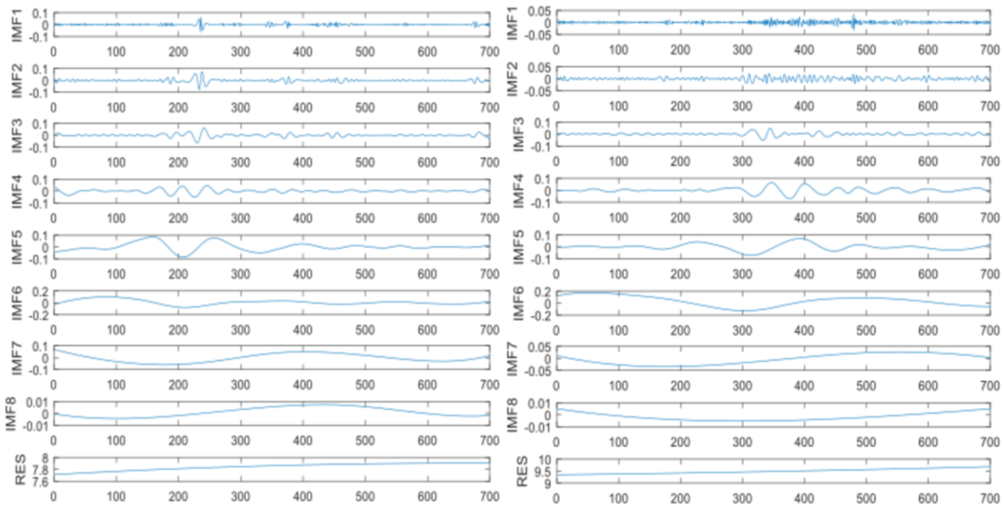


a. vegetable meal

b. soybean meal



c. stalked rice   d. strong wheat



e. early Indica rice   f. Zheng Cotton

Figure 2: Intrinsic Model Functions of Future Prices

4. Forecast Evaluation

We now turn to the main issue of this study, which is to evaluate the optimal combination forecasting performance using the Ensemble Empirical Mode Decomposition (EEMD) technique. Our interest is on short term forecasting. Hence, we only consider one and three days ahead forecasting in this paper. The Neural network, Support Vector Machine and ARIMA models are estimated using the first 80% (560 observation) of the data. Post-sample forecasts for these models and from the optimal combination forecasts are computed for the remaining 20% (140 observations). The post-sample Relative Root Mean Square Errors (RRMSE) give below is used to measure performance.

$$RRMSE = \frac{\sum_{t=560}^{700} (\hat{x}_{t+k} - x_{t+k})^2}{\sum_{t=560}^{700} (\tilde{x}_{t+k} - x_{t+k})^2}$$

Where $\hat{x}_{t+k}$ is the k-step ahead forecast computed by the optimal combination method and $\tilde{x}_{t+k}$ is the k-step ahead forecast obtained either by SVM, NN or ARIMA.

4.1 Optimal Models for Intrinsic Model Function Components

In this paper, radial basis kernel function is selected to find the optimal support vector machine model. We utilize Grid search (Gridsearch), Genetic Algorithm (GA) and Particle Swarm Optimization (PSO) to optimize the parameters. The input vector dimension varies from 1 to 5 and the cross-intersection method is used to optimize the training set. The optimal dimension parameter and algorithm for SVM are chosen by comparing the prediction accuracy.

NEWFF () function is utilized to establish the neural network model, the maximum training times is 5000, the learning efficiency is 0.01 and the training precision is selected as 0.001. Tansig and logsig functions are chosen as the activation function, the trainlm and the traingd are selected as the training functions of neural network. The number of neurons in the hidden layer is set from 2 to 10. The optimal activation function, training function and the number of neurons of each component for six categories is obtained by using cross experiment and comparing their prediction errors.

For each IMF and remainder of agricultural products futures prices for six categories, the unit root and white noise tests are carried out. The BIC criterion is used to find the optimum number of lags and the established model is subjected to parameter estimation and parameter test in turn. The optimal ARIMA model were then employed to compute the forecasts.

Finally, the optimal combination prediction models corresponding to each component is chosen by comparing their prediction error (RMSE) The final optional forecast models are given as Table 2. (Further details about the selection of optimal models and forecasting performance of the decomposed components are available from the authors upon request).

Table 2: The Optimal Model for IMFs for the six categories

	vegetable meal	soybean meal	stalked rice	strong wheat	early Indica rice	Zheng Cotton
IMF1	NN (L,TR,7)	ARIMA (4,0,0)	SVM(G,4)	SVM(G,4)	NN (L, TR,9)	NN (L,TR,10)
IMF2	NN (L,TR,9)	SVM(P,5)	NN (L,TR,5)	NN (L,TR,3)	SVM(GS, 5)	NN (T,TR,5)
IMF3	ARIMA (6,0,0)	ARIMA (6,0,0)	ARIMA (6,0,0)	ARIMA (6,0,0)	ARIMA (6,0,0)	ARIMA (6,0,0)
IMF4	ARIMA (4,0,2)	ARIMA (6,0,0)	ARIMA (6,0,0)	ARIMA (6,2,0)	ARIMA (6,0,0)	ARIMA (6,2,2)
IMF5	SVM(G,2)	SVM(G,5)	ARIMA (6,0,0)	SVM(G,4)	SVM(G,2)	SVM(G,2)
IMF6	ARIMA (6,0,0)	SVM(G,2)	ARIMA (6,1,0)	ARIMA (5,1,0)	ARIMA (6,1,0)	ARIMA (6,1,0)
IMF7	ARIMA (6,0,0)	ARIMA (6,0,0)	SVM(G,4)	ARIMA (5,0,0)	ARIMA (6,0,0)	ARIMA (6,1,0)
IMF8	ARIMA (4,0,0)	ARIMA (1,1,0)	NN (T,TR,7)	NN (L,TR,6)	ARIMA (4,1,0)	NN (T,TR,5)
RE	ARIMA (6,1,0)	SVM(G,2)	NN (L,GS, 2)	NN (T,TR,5)	ARIMA (2,1,0)	NN (L,TR,5)

Note:L indicates logsig function, TR indicates trainlm function, T indicates tansig function, P indicates PSO function, G indicates GA function and GS indicates GridSearch function.

We can generally conclude from Table 2 that the non-linear models (SVM and NN) are more suitable for the high-frequency components, (IMF1 and IMF2), except for IMF1 of soybean meal (AR model). However, for the low frequency components and the remainder, there is no obvious pattern for selection of the individual models.

4.2 Forecast Results

In this paper, we combine the forecast results for all components of each category decomposed by EEMD method and obtain the final combination prediction results for each agricultural commodity futures price. Support vector machine, neural network and ARIMA model are chosen as the benchmark models for comparison. Table 3 presents the post-sample Relative Mean Square Error (RRMSE) for the one-day ahead forecasts of the six agricultural products.

Table 3: One-step ahead post-sample RRMSE

Category	EEMD/SVM	EEMD/NN	EEMD/ARIMA
Vegetable meal	0.6968**	0.7031**	0.6922**
Soybean meal	0.6248**	0.6231**	0.6209**
Stalked rice	0.6144**	0.6241**	0.5136**
Strong wheat	0.6565	0.6558	0.6517
Early Indica rice	0.8272**	0.8324**	0.8256**
Zheng Cotton	0.7796**	0.7755**	0.7738**
Average	**0.7043**	**0.7067**	**0.6872**

Diebold-Mariano test was employed to test the equality of forecast errors between EEMD and the benchmark models for individual product types. *: significant at 5% , **: significant at 1% level

It can be seen from Table 3 that the prediction errors among the two individual models, support vector machine, neural network, are almost the same, and they are slightly better than the linear ARIMA model. The prediction error of the combined model is much smaller than that of support vector machine, neural network and ARIMA model for all the six products, which suggests the superiority of the combined model utilizing EEMD approach. In fact, the combined model outperforms the SVM, NN and ARIMA by 30%, 30% and 32% respectively.

To further test the superiority of the proposed combination approach in this paper, we also computed the three-steps ahead forecasts for the futures realization prices of the six products, with SVM, NN and ARIMA as the benchmark models for comparison. Table 4 shows the post-sample RRMSE.

Table 4: Three-step ahead post sample RRMSE

Category	EEMD/SVM	EEMD/NN	EEMD/ARIMA
Vegetable meal	0.6166**	0.5417**	0.6051
Soybean meal	0.4913*	0.4869*	0.5504*
Stalked rice	0.6684**	0.6774**	0.6819**
Strong wheat	0.6490	0.5796	0.6859
Early Indica rice	0.7505	0.7143	0.8955
Zheng Cotton	0.5024**	0.4764**	0.5024**
Average	**0.6199**	**0.5863**	**0.6657**

Diebold-Mariano test was employed to test the equality of forecast errors between EEMD and the benchmark models for individual product types. *: significant at 5% , **: significant at 1% level

We employed the Diebold-Mariano test to test the significance of 1-step ahead and 3-steps ahead forecasting errors between EEMD and SVM, NN, ARIMA models, and reported the result in Table 3 and Table 4. For one-step ahead forecast error tests, we arrived the conclusions that the errors of EEMDs are at 1% significantly less than the three benchmark models for all commodities. For statistical tests of three-steps ahead post sample forecast errors, we concluded that for all commodities , the results are significant either at 1% or 5% level. Only for Strong Wheat and Early Indica Rice, the forecast errors are not significantly different.

From Table 4, again the results support the superiority of the combined models with EEMD in all cases. In fact, for the 3 days ahead forecasts, the gains in the accuracy are more than the gains obtained for the 1 day ahead forecasts. They outperform the SVM, NN and ARIMA models by 38%, 42% and 33% respectively. The results for the hybrid non-linear models are generally better than the linear ARIMA.

5. Conclusions

In this study, the futures prices of 6 categories of vegetable meal, soybean meal, stem rice, strong wheat, early Indica rice and Zheng cotton were decomposed by utilizing the ensemble empirical mode decomposition approach. The combination models of support vector machine, neural network and ARIMA model were then used to predict the agricultural futures prices of this six categories.

Comparing the combined models with the benchmark models, SVM, NN and ARIMA, showed that the prediction performance of the combination models is superior to that of individual models. With the increase of the prediction horizons, the superiority of the combined models using the empirical mode decomposition (EEMD) becomes more pronounced. The performance of the two non-linear models are better than the linear ARIMA, however, there is no obvious difference between the two non-linear models. In particular, the results indicate higher accuracy in forecasting high frequency components using SVM and Neural network than that of ARIMA models, which show that support vector machine and Neural Network are more suitable for the predicting high frequency components.

Acknowledgements

The research was partially supported by a financial grant from the National Natural Science Foundation of China (Project 71971089). The authors are grateful to the editor and associate editor for constructive comments.

References:

Baruník, J., Malinská, B. 2016. Forecasting the term structure of crude oil futures prices with neural networks. Applied Energy, 164: 366-379.

Bohl, M. T., Siklos, P. L., Wellenreuther, C., 2018. Speculative activity and returns volatility of Chinese agricultural commodity futures. Journal of Asian Economics, 54: 69-91.

Bonato, M., Demirer, R., Gupta, R., Pierdzioch, C. 2018. Gold futures returns and realized moments: A forecasting experiment using a quantile-boosting approach. Resources Policy.

Box,G, Jenkins,G., 1970. Time Series Analysis: Forecasting and Control. Holden-Day, San Francisco.

Byun, S. J., Cho. H., 2013. Forecasting carbon futures volatility using GARCH models with energy volatilities. Energy Economics, 40: 207-221.

Cartwright, P. A., Riabko.,N. 2015. Measuring the effect of oil prices on wheat futures prices. Research in International Business and Finance, 33: 355-369.

Chen, Lin, Huang, Zhangshu, 2010. Futures Price Analysis and Forecast Based on ARIMA Model. Journal of Fuzhou University (Philosophy and Social Sciences),97(3):32-37

Chu, Dong, 2014. Study on Comparative Prediction of Soybean Futures Price by Time Series Model. Finance Economy, (24):116-119

Chu, H.-H., Chen , T.-L., Cheng, C.-H., Huang, C.-C., 2009. Fuzzy dual-factor time-series for stock index forecasting. Expert Systems with Applications, 36 (1): 165-171.

Fang, L., Yu, H., Xiao, W., 2018. Forecasting gold futures market volatility using macroeconomic variables in the United States. Economic Modelling.

Fang, L., Yu, H., Xiao, W., 2018. Forecasting gold futures market volatility using macroeconomic variables in the United States. Economic Modelling.

Fileccia, G., Sgarra, C. 2018. A particle filtering approach to oil futures price calibration and forecasting. Journal of Commodity Markets, 9: 21-34.

Gao, Xinyu, Yu, Guoxin, 2014.. Research on the Method of Forecasting Cotton Futures Price in China Based on Comparison between EGARCH-EWMA and ARIMA model. Price: Theory and Practice,366(12):85-87

Hamid, S. A., Iqbal., Z., 2004 Using neural networks for forecasting volatility of S&P 500 Index futures prices. Journal of Business Research, 57 (10): 1116-1125.

Hamid, Shaikh A, Iqbal, Zahid, 2004. Using neural networks for forecasting volatility of S&P 500 Index futures prices. Journal of Business Research, 2004,57(10): 1116-1125

Huang, Wen, Huang, Zhuo, Wang, Tianyi, 2013. Volatility and Correlation prediction of high Frequency Agricultural Futures: Based on Realized Copula-DCC Model. Zhejiang Social Sciences, (05): 40-47+156

Kang, S. H., Yoon, S.-M. 2013. Modeling and forecasting the volatility of petroleum futures prices. Energy Economics, 36: 354-362.

Li, G., Li., Y., 2015. Forecasting copper futures volatility under model uncertainty. Resources Policy, 46: 167-176.

Li, Z., Lu., X., 2011. Multifractal analysis of China's agricultural commodity futures markets. Energy Procedia, 5: 1920-1926.

Liang, Jingxi, Tai, Yinping, 2014. Price forecasting of China's soybean futures based on the EEGARCH-EWMA model. Science Technology and Management, 2014,16(2):58-62

Lin, Jie, Gong, Zheng, 2017. A Research on Forecasting of Shanghai Zinc Futures Price Based on Artificial Neural Network. The Theory and Practice of Finance and Economics, 38(2):54-57

Liu, J., F. Ma, K. Yang, Y. Zhang, 2018. Forecasting the oil futures price volatility: Large jumps and small jumps. Energy Economics, 72: 321-330.

Luo, X., Ye., Z., 2015. Predicting volatility of the Shanghai silver futures market: What is the role of the U.S. options market? Finance Research Letters, 15: 68-77.

Ma, F., J. Liu, D. Huang, W. Chen, 2017. Forecasting the oil futures price volatility: A new approach. Economic Modelling, 64: 560-566.

Rumelhart, D. E, McClleland, J. L, 1986. Learning internal representation by error propagation. Parallel Distributed Processing: Exploration in the Microstructure of Cognition,1(8):318-362.

Sévi, B., 2014. Forecasting the volatility of crude oil futures using intraday data. European Journal of Operational Research, 235 (3): 643-659.

Teng, Yongping; Zhou, Tingting, 2017. Empirical study on price prediction of soybean futures. Journal of Shenyang University of Technology (Social Science Edition), 2017,10(4):331-334

Tian, F., Yang, K., Chen, L., 2017. Realized volatility forecasting of agricultural commodity futures using the HAR model with time-varying sparsity. International Journal of Forecasting, 33 (1): 132-152.

Tian, Fengping, Yang, Ke., 2016. Forecasting Realized Volatility of Algricultural Commodity Futures using TVS-HAR Model. Systems Engineering-Theory and Practice.(12): 3003-3016

Vapnik, V.,1998．Statistical learning theory. New York, NY：Wiley, Chapter 10, 401-405

Wang, Zicheng, 2015. Prediction of Carbon Futures Price Based on ARIMA-LSSVM Model. Harbin: Harbin Institute of Technology

Wei, Ruiduo, 2009. The forecast Model of the Spot Price Based on Future Price-- with the Example of Copper Future in SHFE. Forecasting, 28(03):61-64.

Wei, Y, 2012. Forecasting volatility of fuel oil futures in China: GARCH-type, SV or realized volatility models? Physica A: Statistical Mechanics and its Applications, 391 (22): 5546-5556.

Wen, F., X. Gong, S. Cai, 2016. Forecasting the volatility of crude oil futures using HAR-type models with structural breaks. Energy Economics, 59: 400-413.

Wu, Z., Huang, N.E.,2009. Ensemble empirical mode decomposition: A noise-assisted data analysis method. Advances in Adaptive Data Analysis,2009,1:1-41.

Xiong, T., C. Li, Y. Bao, Z. Hu, L. Zhang,2015. A combination method for interval forecasting of agricultural commodity futures prices. Knowledge-Based Systems, 77: 92-102.

Yang Jianhui, Dai, Xiaozhen, 2013. International Nonferrous Metals Futures Price Forecast Based on Improved FSA-SVM. Statistics and Decision, 2013,374(2):165-167

Yang, C., X. Gong, H. Zhang, 2018. Volatility forecasting of crude oil futures: The role of investor sentiment and leverage effect. Resources Policy.

Yang, Chao, Liu, Xihua, 2014. Forecasting Stock Index Futures Price Based on Wavelet and BP Network. Journal of Qingdao University (Natural Science Edition),1(27):101-

105

Yang, K., Tian, F., Chen, L., Li, S..,2017. Realized volatility forecast of agricultural futures using the HAR models with bagging and combination approaches. International Review of Economics & Finance, 49: 276-291.

Yang, Ke, Tian, Fengping, 2014a. Realized Volatility Dynamics and Forecasting Model for Agricultural Commodity Futures. Economic Review, 2014, (04):50-67

Yang, Ke, Tian, Fengping, 2014b. Volatility Projection for Agricultural Commodity Futures under Structural Breaks. Acta Scientiarum Naturalium Universitatis Sunyatseni,2014, (02):59-72

Yeh, Jia-Rong, Sun ,Wei-Zen, Shieh, Jiann-Shing, Huang, Norden E., 2010. Medical Engineering & Physics. Medical Engineering & Physics, 32:490-496

Yousefi, Shahriar, Weinreich, Ilona, Reinarz., Dominik, 2005. Wavelet-based prediction of oil prices. Chaos Solitons and Fractals, 25(2):265-275

Zhang, Xin., 2012 A Study of Corn Futures Prices Prediction Based on Chaotic Time Series. Shenyang: Northeast Agricultural University

Data Availability Statement

The data that support the findings of this study are available in Wind at [www.wind.com.cn]. The Wind database is subscription-based. Six categories of agricultural futures are selected from the database: vegetable meal, soybean meal, stalked rice, strong wheat, Zheng Cotton and early Indica rice. The data of futures prices are the daily closing prices, and in all cases, our sample period starts from 27th November 2014 and ends in 11th October 2017. All six product prices have the same reference number 1711.