

The Empirical Analysis for the Spread of Soya Oil and Soybean Meal Based on Wavelet Neural Network

Hao-Ting Li¹, Xiao-Jie Liu², Yuan-Biao Zhang^{3,4}, Ya-Hao Fu¹ & Jian-Yu Zheng⁵

¹ International Business School, Jinan University, Zhuhai, China

² Electrical and Information School, Jinan University, Zhuhai, China

³ Packaging Engineering Institute, Jinan University, Zhuhai, China

⁴ Key Laboratory of Product Packaging and Logistics of Guangdong Higher Education Institutes, Jinan University, Zhuhai, China

⁵ School of Translation Studies, Jinan University, Zhuhai, China

Correspondence: Hao-Ting Li, International Business School, Jinan University, Zhuhai 519070, China. E-mail: sainianjian@outlook.com

Received: March 16, 2015

Accepted: April 15, 2015

Online Published: May 25, 2015

doi:10.5539/ijef.v7n6p80

URL: <http://dx.doi.org/10.5539/ijef.v7n6p80>

Abstract

For the sake of a better cross-commodity arbitrage in the futures market, WNN (wavelet neural network) is adopted to analyze the previous spread and predict the future in this paper. Firstly, the correlation coefficient of previous prices between the two goods is calculated in order to examine whether there is arbitrage opportunity. Considered that the spread could be affected by many nonlinearity factors and BP neural network has slow convergence rate, BP neural network is combined with wavelet analysis which has excellent partial analysis ability. In this way, the prediction model about soya oil and soybean meal spreads is built based on WNN. Compared the result calculated through that method with only BP neural network's: WNN is superior to neural network in predicting rapid fluctuation and secular trend.

Keywords: spread-prediction, cross-commodity arbitrage, WNN, BP neural network

1. Introduction

Cross-commodity arbitrage is a means of exchange through the unreasonable price difference between futures in order to make venture profits. Tang (2006) summarized various types of spread arbitrage, including intertemporal arbitrage, cross-market arbitrage and cross-commodity arbitrage. With the continuous development of China's commodity futures market, cross-commodity arbitrage is appealing to investors, after calendar spread arbitrage and cross market arbitrage (Li, 2010).

Ma (1988) had a research on gold's and silver's spot and futures market, finding that price have consistency in the two stages, where there are short-term arbitrage opportunities caused by spread. Tse and Booth (1995) studied the relationship between Treasury bonds futures and eurodollar futures, and the result is that there is a cointegration relationship between the two futures markets. Tian (2013) used market data to prove that there existed arbitrage opportunities relying on soybean crush, while he also noted that the market and trading risk should not be ignored owing to the immature domestic futures market. Effective analysis and trend prediction are important means of controlling price risk and getting revenues, which makes the domestic market continue to mature.

Moving average method, exponential smoothing method and grey prediction are generally used to predict price changes in futures market. Due to nonlinear factors, such as economic policy and investor psychology, which could affect futures price, the results of traditional prediction method are unsatisfactory. Neural network, with strong nonlinear approximation, can deal with nonlinear relationship between input variables and output variables better (Wang, 2009). Grudnitski and Osburn (1993) used neural network to predict the spread between S&P index and gold price. Shaikh A. Hamid and Zahid Iqbal (2004) used neural network to predict the S&P 500 index futures price. Neural network is likely to be caught into a local optimum in the learning process, so wavelet analysis with strong local analytical capacity is introduced to improve neural network, which makes the

new model possess the ability of flexible and effective approximate function and fault tolerance (Zhang, Yu, & Li, 2010). In this paper, WNN was employed to forecast spread, compared with the prediction result of BP neural network in the short term and the long term.

2. Spread-Prediction Model

The correlation coefficient between one good and another in price can reflect the fluctuation relationship of each price (Lu, 2011), which is crucial to cross commodity arbitrage. The prediction of spread can contribute to investors' decision-making, so the model where the spread is predicted through the past data is proposed. And then this paper have reasoned the application feasibility of WNN to the spread.

2.1 Correlation Test

Cross-commodity arbitrage of commodity futures affects the contract price in futures market according to the co-movement in market demand and supply, which causes strong correlation between different commodities in future price. So the price correlation between commodity future through correlation coefficient is analyzed to judge the arbitrage feasibility between commodity futures. The correlation coefficient is calculated as follows:

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}} \quad (1)$$

Where, x_i, y_i are the future price series of relevant commodities respectively, \bar{x}, \bar{y} are the average price respectively, r is the correlation coefficient between the futures price trend of relevant commodities. As to cross-commodity arbitrage in commodity futures, it doesn't mean that it is better if the correlation is higher, for example, when the relation is 1, the arbitrage chance is lost because the tendency of two commodities is the same. By looking up relevant books, when the correlation coefficient is between 0.7 and 0.85, the arbitrage effect is fine.

2.2 Wavelet Neural Network

BP neural network, put forward in 1986, is a multi-layer and feed-forward neural network, which trains data in accordance with error back-propagation algorithm. Through Continually learning and storing large amounts of information, BP neural network forms a mapping relation model with a large number of input-output, and weights and thresholds are improved and amended constantly during training. Because of slow convergence rate, scholars would like to introduce other methods to offset the shortcoming and improve the prediction accuracy.

Wavelet analysis is a method of time-frequency analysis according to signals, where, in the fixed window, signals can be partially analyzed through changing their waves, which makes them have fine frequency resolution in the low frequency part and fine temporal resolution in the high frequency part. When wavelet analysis is introduced into BP neural network, wavelet neural network(WNN) can take full use of strong local learning ability of wavelet analysis and self learning ability of BP neural network, to make it possess a better ability of approximation and fault tolerance. Based on topological structure of neural network, WNN takes it as the transfer function of the hidden layer. The structure is as followed:

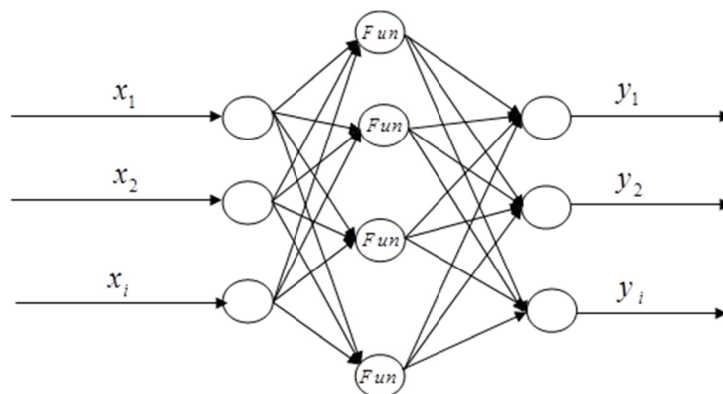


Figure 1. The topological structure of WNN

Where, x_1, x_2, \dots, x_i are the input parameters of wavelet neural network, y is the predicted output value, Fun is wavelet basis function, w_{ij} and w_{jk} are the weight of wavelet neural network. when the input signal's series is $x_i (i = 1, 2, \dots, k)$, the calculate equation for hidden layer is:

$$h(j) = h_j \left(\frac{\sum_{i=1}^k w_{ij} x_i - b_j}{a_j} \right) \quad j = 1, 2, \dots, l \quad (2)$$

Where, $h(j)$ is the output value in hidden layer's j th nodes; w_{ij} is connected weight between input layer and hidden layer; b_i is the horizontal shift factor of wavelet basis function h_i ; a_j is contraction-expansion factor of wavelet basis function h_j . The wavelet basis function used in this model is Morlet.

$$y = \cos(1.75x) e^{x^2/2} \quad (3)$$

The calculate equation of wavelet neural network's output layer is:

$$y(k) = \sum_{i=1}^l w_{ik} h(i) \quad k = 1, 2, \dots, m \quad (4)$$

Where, w_{ik} is the weight from hidden layer to output layer; $h(i)$ is the output of the i th hidden layer node; l is the node number in hidden layer; m is the node number in output layer.

In the process of modification, the weight need be modified according to the error of WNN when the weight meets the requirement of wavelet neural network. Prediction error $\Delta y(k)$ is calculated by:

$$e = \sum_{k=1}^m yn(k) - y(k) \quad (5)$$

Where, $yn(k)$ is the expected output value; $y(k)$ is the forecast value of wavelet neural network. The modify the weight of wavelet neural network and the coefficient of wavelet neural network according to e obtained.

$$w_{n,k}^{i+1} = w_{n,k}^i - \eta \frac{\partial e}{\partial w_{n,k}^i} \quad (6)$$

$$a_k^{i+1} = a_k^i - \eta \frac{\partial e}{\partial a_k^i} \quad (7)$$

$$b_k^{i+1} = b_k^i - \eta \frac{\partial e}{\partial b_k^i} \quad (8)$$

Where, η is learning velocity. The training process is as followed:

Step1: define the contraction-expansion factor a and the shift factor b of wavelet neural network randomly, and set connected weight w_{ij} , w_{jk} and the learning velocity η . Then initialize all of the grid.

Step2: divide the sample into a training set and a test set. The training set is used to train the network, and the test set is used to analyze the forecast result of the grid.

Step3: obtain the forecast result and output, and calculate the deviation between them.

Step4: decide whether it is necessary to re-modify the weights and parameters according to deviation. And decrease deviation constantly and let it close to actual value until we find the best forecast result (MATLAB Chinese Forum, 2010).

3. Empirical Analysis for Spread Prediction of Soya Oil and Soybean Meal

Soya oil and soybean meal are two kinds of main products after soybeans are crushed, belonging to the same industry chain, so there are arbitrage opportunities in the futures market. Through the correlation analysis from historical datum, we can judge the possibility for arbitrage. WNN is adopted to predict the spread between soya oil and soybean meal, and then the spread trend and predict outcomes should be objective analysed so as to make profits from the spread. The data of soya oil futures and soybean meal futures is from Dalian Commodity Exchange between 2009 and 2012 (total 973 workdays), The result by WNN is compared with that by neural

network in long term and short term.

3.1 The Correlation Coefficient between Both Futures Prices

Soya oil and soybean meal belong to the same industry chain, so there is co-movement between both futures prices to some extent. Pearson Correlation Coefficient is engaged in quantitative analysis of relationship between soya oil futures prices and soybean meal futures prices:

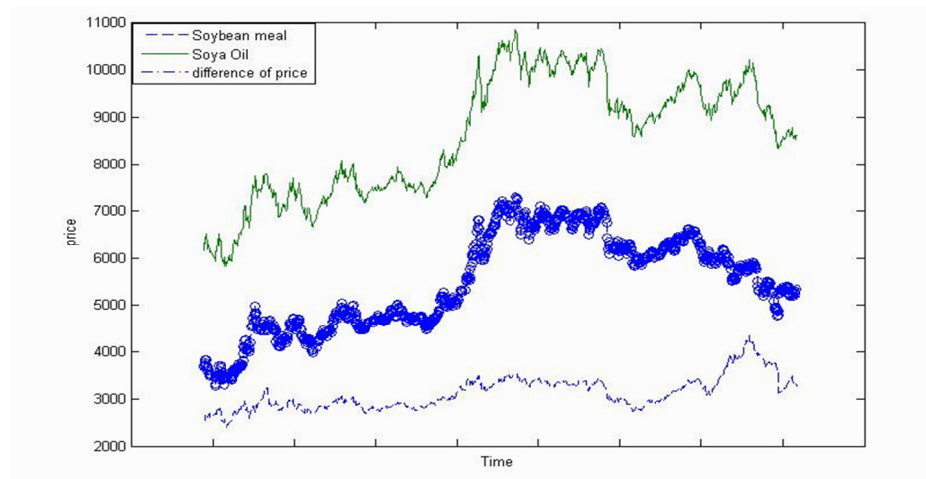


Figure 2. The price trend of soya oil and soybean meal

In Figure 2, the solid line above stands for soya oil futures prices, and the dotted line below stands for soybean meal futures prices. Although the volatility of the two futures prices is different (volatility of soya oil is greater than that of soybean meal), their price trends are similar. Apparently the fluctuation range of spread is broad, with the minimum value at about 3000 yuan and maximum value at around 7000 yuan, which provides many arbitrage opportunity.

Table 1. The correlation coefficient

		Soybean Meal	Soya Oil
Soybean Meal	Pearson correlation	1	0.767**
	Significant(two-sided)		0.000
	N	973	973
Soya Oil	Pearson correlation	0.767**	1
	Significant(two-sided))	0.000	
	N	973	973

** .01 level (bilateral) is significantly related.

Table 1 shows the Correlation Coefficient of both futures prices is 0.767 based on the 973 future prices in these trading days of the past four years, meeting the range of the coefficient coefficient, which is feasible to guarantee the security and feasibility of spread arbitrage. To gain profits from price fluctuation and to control price risk, it is vital to predict the spread accurately (Lu, 2011).

3.2 The Spread Prediction Based on WNN

There is a database about futures prices of soya oil and soybean meal, where it contains 973 elements each of which represent a relevant work-day datum. WNN and BP neural network are employed to forecast spread in both short term and long term. Due to the difference of the prediction length, raw elements in the database form respectively short-term prediction set whose length is 972 and long-term prediction sets whose length is 953. Short-term prediction means forecasting the next workday's spread based on historical elements, and the first 922-workday elements are used as a training set, then we train these elements by WNN. The next 50 elements will be used as a test set to test predictive ability of this model; On the other hand, the long-term prediction means forecasting the next 20 workdays' spread. The first 933 elements are considered as a training set, the rest

is used as a test set in the long-term prediction. The number of input nodes is 12, namely opening price, high price, low price, closing price, volume and open interest for each; and output node number is 1: the second day's spread; number of hidden nodes is 17.

Table 2. Training set and test set

		Training set				Test set		
Soya oil	Opening price	6220	6300	6600	...	8606	8562	8620
	High price	6288	6520	6726	...	8650	8636	8674
	Low price	6160	6226	6442	...	8542	8524	8610
	Closing price	6240	6470	6490	...	8572	8634	8642
	volume	554070	701442	680826	...	552382	692036	507182
	Open interest	154364	159354	142176	...	547060	537370	507428
Soybean meal	Opening price	2500	2560	2628	...	3316	3315	3362
	High price	2574	2632	2674	...	3358	3370	3374
	Low price	2494	2525	2616	...	3309	3306	3332
	Closing price	2568	2609	2621	...	3333	3368	3336
	volume	578608	778132	751362	...	2280378	2155590	1649790
	Open interest	209878	217790	210822	...	1270114	1391796	1257114
spread		3672	3861	3869	...	5239	5266	5306

Table 2 provides concrete datum, such as opening price and closing price of soya oil and soybean meal. In order to avoid effects of different dimensions and different ranges of fluctuations, datum should be normalized firstly. The dimension of input vectors is 12, including opening price, high price, low price, closing price, volume and open interest of soya oil and soybean meal respectively; The dimension of output vector is 1, namely is the spread. The datum in the front belong to the train set and the latter datum belong to the test set in Table 2. Then we compare results with the results by BP neural network. [4] The test is carried out based on WNN and neural network 200 times for both short-term prediction and long-term prediction. One of results is shown in Figure 2 and Figure 3.

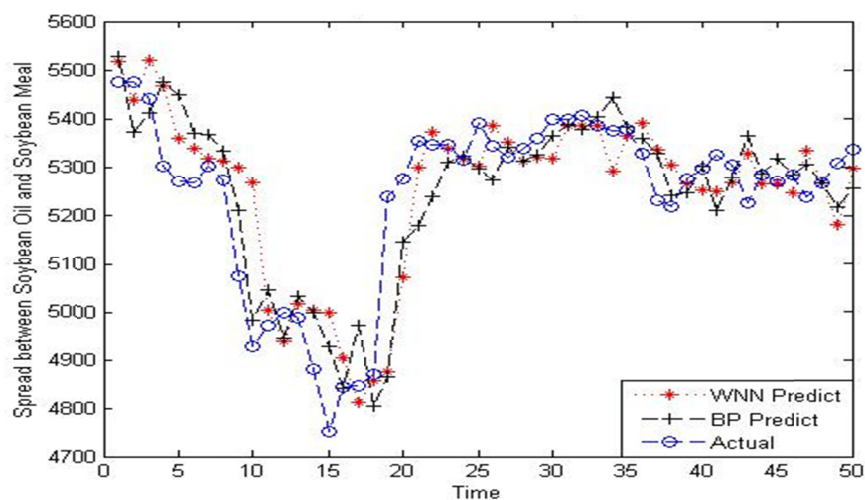


Figure 2. Short-term spread-prediction

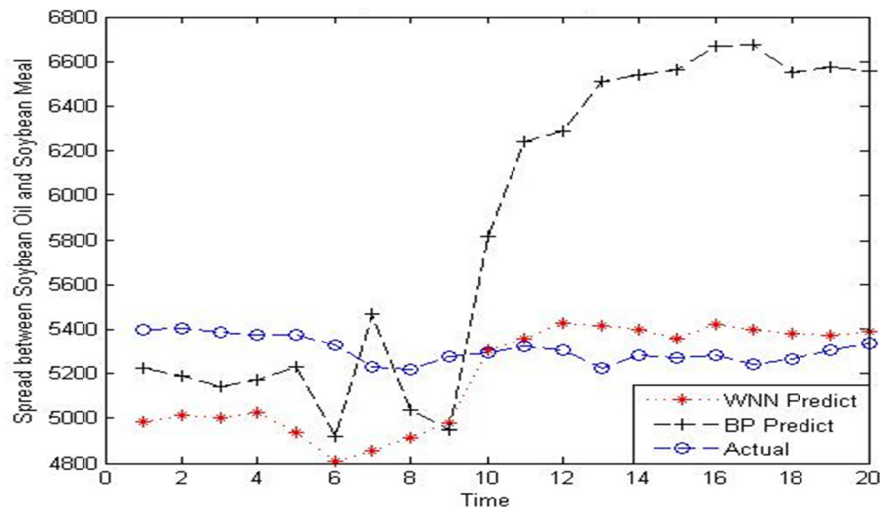


Figure 3. Long-term spread-prediction

Figure 2 shows that WNN and BP neural network can predict the spreads with high precision in the short term. In Figure 3, WNN and BP neural network do poor in long-term prediction, although in the field of trend prediction, the prediction results by WNN is better. The fluctuation predicted by WNN is smoother than the actual one, while that by BP neural network is greater than the actual one. To analyze the prediction results precisely, the following indicators are adopted: mean absolute percentage error (MAPE) and root mean squared error (RMSE) (Li, Xu, & Chen, 2009):

$$MAPE = \frac{1}{N} \sum \left(\frac{1}{n} \sum \left| \frac{f_i - \hat{f}_i}{f_i} \right| \right) \quad (9)$$

$$RMSE = \frac{1}{N} \sum \left[\frac{1}{n} \sum \left(\left| \frac{f_i - \hat{f}_i}{f_i} \right| - \left(\frac{1}{n} \sum \left| \frac{f_i - \hat{f}_i}{f_i} \right| \right) \right)^2 \right]^{\frac{1}{2}} \quad (10)$$

Where, f_i is the i th actual value in the training set, and \hat{f}_i represents the i th predicted value. The results are given in Table 2.

Table 3. Results of indicators

	Short-term prediction		Long-term prediction	
	MAPE	RMSE	MAPE	RMSE
WNN	0.0013	0.0133	0.0188	0.0466
BPNN	0.0095	0.0149	0.1408	0.1495

Table 3 shows that, in the short-term prediction, MAPEs of both WNN and BP neural network are under 1%, which means each forecast accuracy is as high as 99%, with MAPE of WNN at 0.0013, less than that of BP neural network at 0.0095; RMSE of WNN is 0.0133, also less than that of BP neural networks (0.0149). Results of WNN is better, the majorization where wavelet analysis is used in BP neural network is effective, with increase by 0.83% in precision. In the long-term prediction, MAPE and RMSE of WNN are significantly smaller than those of BP neural network. The precision is increased by 14.19%. The result in Figure 3 also shows that WNN predicts the trend, more accurately than BP neural network does. WNN performs better than BP neural network in both short-term and long-term prediction.

4. Conclusion

In order to grasp market better, to guarantee the feasibility and safety, and to increase the possibility of cross-commodity arbitrage, WNN is employed to forecast the spread, guiding operation of spread-arbitrage, then take the spread between soya oil and soybean meal for the example.

Firstly, by analyzing the correlation between soya oil and soybean meal, arbitrage opportunity can be figured out. According to opening prices, high prices, low prices, closing prices, volumes and open interests, WNN and BP neural network are employed to predict the spread in the long term and short term. Then we compared the difference in precision between the two different methods. The result shows that WNN is superior to neural network in the process of long-term and short-term predictions.

Although WNN is superior to neural network, it is necessary to improve accuracy of long-term prediction. because the number of factors that affects spread is more than that of individual commodity, and the spread is predicted merely according on opening price, high price low price and so on that, the information is not taken full use of. For sake of accurate prediction, more factors should be taken into consideration, such as import and export of related goods and interest rate. In addition, the prediction model should be connected with reality, so that the accuracy of it would be higher.

Reference

- Christopher, K. M., & Luc, A. S. (1988). Arbitrage Opportunities in Metal Futures Markets. *The Journal of Futures Markets*, 8(2) 199-209. New York, NY. <http://dx.doi.org/10.1002/fut.3990080207>
- Grudnitski, G., & Osburn, L. (1993). Forecasting S&P and Gold Futures Prices: An Application of Neural Networks. *The Journal of Futures Markets*, 13, 631-643. <http://dx.doi.org/10.1002/fut.3990130605>
- Li, D., Xu, W., & Chen, R. (2009). Real Estate Price Forecast Using Rough Sets and Wavelet Neural Networks. *Management Review*, 11, 18-22.
- Li, L. (2010). *Study for Risk of Cross-species Arbitrage on Commodity Futures*. Chengdu, China: Southwestern University of Finance and Economics.
- Lu, Y. (2011). *Research on Soybean Oil and Palm Oil Arbitrage Trading Model*. Chengdu, China: Southwest Jiaotong University.
- MATLAB Chinese Forum. (2010). *MATLAB Neural Network Analysis of 30 Case* (pp. 208-210). Beijing, China: Beihang University Press.
- Shaikh, A., & Hamid, Z. I. (2004). Using neural net works for forecasting volatility of S&P 500 Index futures prices. *Journal of Business Research*, 57(10), 1116-1125. [http://dx.doi.org/10.1016/S0148-2963\(03\)00043-2](http://dx.doi.org/10.1016/S0148-2963(03)00043-2)
- Tang, Y. (2006). *Study on Theory and Application of Commodity Futures Spread Trading*. Shanghai, China: Tongji University.
- Tian, Y. (2013). *Empirical Studies of Cross-commodity Arbitrage in Futures Market*. Chengdu, China: Southwestern University of Finance and Economics.
- Tse, Y., & Booth, G. (1995). the Relationship between U.S. and Eurodollar Interest Rates: Evidence from the futures markets. *Weltwirtschaftliches Archiv*, 13(1), 28-46. <http://dx.doi.org/10.1007/BF02709070>
- Wang, H. (2009). *The Method of Futures Prices Forecasting Research Based on Neural Network*. Beijing, China: Capital Normal University.
- Zhang, K., Yu, Y., & Li, T. (2010). Application of Wavelet Neural Network in Prediction of Gold Price. *Computer Engineering and Application*, 27, 224-226.

Copyrights

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/3.0/>).