Prediction of Futures Price Fluctuation: Modeling, Computing and Analysis System

Weihui Dai  
School of Management, Fudan University, Shanghai 200433, P.R.China  
Email: whdai@fudan.edu.cn

Xingyun Dai, Hua Ye and Jie Liu  
School of Management, Fudan University, Shanghai 200433, P.R.China  
Email: {082025041, 0525005, liujie}@fudan.edu.cn

Abstract—With the development of global market and information network, international futures markets have become integrated. Price fluctuation of one futures market may cause related price fluctuation in another exchange. Many researches have shown that there is high correlation between international and domestic futures price fluctuation. However, quantitative research and the information system for analyzing and forecasting the fluctuation based on that correlation are still expected to be further explored.

This paper studied the price fluctuation and its correlation between two different futures markets, and presented a quantitative model for the prediction in a following market by using spectrum analysis of time series. Based on those researches, an analysis system was developed on .net platform to realize the prediction of price fluctuation. In this system, we explored the computing technology with mathematical programming language R, and integrated it with .net platform, thereof making the prediction calculation precise and convenient. Test by this system has shown the accurate result of prediction in the guided market.

Index Terms—international futures, prediction model, correlation analysis, analysis system, computing technology

I. INTRODUCTION

As a type of financial instrument, futures provide the buyers a way to purchase the underlying commodity for a specific price across time. There are two kinds of futures, commodity futures and financial futures. With futures, hedgers can evade risk. For example, a farmer may know the price of corn in autumn before planting in spring. Therefore, he can forecast the demand and make decision for the supply, according to the price in future.

The agreement of the transaction is futures contract. Notice that futures contracts can be sold and bought every trading day, those who buy futures contracts often sell the futures contracts they bought before maturity. For instance, if you believe the corn price of next year is underestimated, you can buy some corn futures contracts and sell them when the contracts price increasing before the delivery date. In this way, the speculators take the risk and make profit with the price fluctuation of futures.

Futures contracts are standardized and mostly traded in exchanges. Influenced by economic globalization, recombination among exchanges occurs frequently. The most famous exchanges by now are Chicago Board of Trade (CBOT), Chicago Merchantile Exchange (CME), Chicago Board Options Exchange (CBOE) and New York Merchantile Exchange (NYMEX) in America, EUREX, Euronext and London Metal Exchange (LME) in Europe. With the development of futures, Chinese futures markets composed of 4 exchanges: Dalian Commodity Exchange (DCX), Zhengzhou Commodity Exchange (ZCX), Shanghai Futures Exchange (SHFX) and China Financial Futures Exchange (CFFEX). Futures markets in China are playing a more and more important role in the international futures markets. SHFE is the third largest exchange for copper futures following LME and NYMEX and DCE becomes the second largest exchange for soybean futures.

With the development of global market and information network, international futures markets have become integrated. Price fluctuation of one futures market may cause related price fluctuation in another exchange. Because each futures market has its own business time, study on the correlation among asynchronous futures markets is helpful for fluctuation prediction in the following market.

II. LITERATURE REVIEW

Nowadays, a lot of studies have been concentrated in the price fluctuation pattern of a certain futures, the correlation between futures and spot price and the price of a certain futures in different markets.

Western scholars are inclined to find the reasons for futures price fluctuation from the view of macroeconomics. Williams and Orice M. [1] analyzed the relationship between energy market and energy future, Christos Floros and Dimitrios V. Vougas [2], Razak M. A. and Bacha O. I. [3] studied the interrelations between the stock index futures and stock markets. Moreover, Daal, Elton; Farhat, Joseph and Wei, Peihwang P. [4] focused on the characteristic of futures itself which could be the reason for price fluctuation. Some scholars have been also interested in forecasting futures price fluctuation. Michael D. Mackenzie and Heather Mitchell [5], Hyndman and Cody B. [6] studies the methods of futures...
price fluctuation forecasting by using ARCH and Gauss model respectively.

Meanwhile, Chinese scholars are more likely to use time series analysis methods to study the relationship among the futures markets, or between futures and spot price. Wang J. [7] used time series and causal analysis methods to study the guiding role of futures price in different futures markets. Liu Deng X and Ma B. [8] analyzed the price fluctuation of different financial futures. Besides, Ye W. [9] and Zhang T. [10] focused on the correlation of the metal futures price fluctuation between LME and SHFE, which is also our concern.

Many of these researches have shown that there is high correlation between international and domestic futures price fluctuation. However, quantitative research and the information system for analyzing and forecasting the futures price fluctuation are expected to be further explored. Our paper is aimed to analyze the price fluctuations of the same kind of futures in international and domestic futures markets, find out the correlation model and guiding role between them and finally develop an analysis system to realize the futures price fluctuation analysis and prediction in the guided market.

III. CORRELATION MODEL

A. Object

As we want to analyze the correlation between the futures price in two different futures markets, the price data is essential. Only the price data in the same contract period will be chosen since a certain kind of futures with different contract periods have different price in the same futures bourse. Furthermore, mismatched data will be excluded. For example, the bourses in America are closed on Christmas just like the Spring Festival in China. Finally, all data must be converted to the same currency unit.

After choosing futures price data, sequence of methods are taken to handle the price data and finally get the price prediction of the following futures market.

B. Data Validity

Before analyzing the correlation between two time series, it is necessary to ensure the validity of the data. Therefore, stationary, co-integration and causality tests are needed here.

1) Stationary Test

ADF (Augmented Dickey-Fuller) is a test for unit root in time series. It is an augmented version of Dickey-Fuller test with more complicated time series. Here we use it to test the stationarity of copper futures price time series. ADF statistic is supposed to be a negative number. The more negative it is, the stronger belief of rejecting the hypothesis that there is a unit root in the time series sample. Only if the time series or their first differences are stationary, the v conditions for co-integration testing are fulfilled.

2) Co-integration Test

Co-integration means that there is real correlation among variables, but not spurious regression. Its basic idea is whether two or more time series are stationary, as long as their linear combination becomes stationary, there is long-run stable relationship among them. Only time series which are integrated of the same order, there may be co-integrated relationship among them. Time series with co-integrated relationship will remain the long-run equilibrium relationship, although noise may exist in short run.

Choose Johansen co-integration test to analyze co-integration relationship between two price series and use the best lag coefficient according to AIC criterion.

If there is co-integration relationship between two series, the relationship can be represented by using ECM (Error Correction Model) as follows.

\[ \Delta CF_i = a_{10} + a_{11} Z_{t-1} + \sum_{j=1}^{p} a_{1j} \Delta CF_{i-j} + \sum_{j=1}^{p} a_{2j} \Delta WF_{i-j} + \varepsilon_{1t} \]

\[ \Delta CF_i = a_{20} + a_{21} Z_{t-1} + \sum_{j=1}^{p} a_{2j} \Delta CF_{i-j} + \sum_{j=1}^{p} a_{3j} \Delta WF_{i-j} + \varepsilon_{2t} \]

CF and WF are the first difference for the two series respectively. Z represents the coefficient of ECM and p is lag coefficient.

3) Granger Causality Test

Co-integrated test tells us whether there is co-integrated relationship between two price series, but Granger causality test is useful to determine whether one time series can be used to forecast another. Here we can use Granger causality test to find the guiding market in the markets.

Because of the sensitivity of lag coefficient, we should use both AIC and SC criterions to get the best lag coefficient.

For further discussion, we study the method for forecasting the price of a certain kind of futures in guided market with the data from guiding market, using cross spectrum analysis.

C. Prediction Model

Basic idea of Spectrum analysis and cross spectrum analysis is that time series are considered as the superposition of sine and cosine waves. Fluctuation pattern of time series can be revealed by comparing variances and circles of the waves. The most useful methods of spectrum analysis include power spectrum analysis and maximum entropy spectrum analysis. Only power spectrum analysis is used in our research.

Time series for spectrum analysis should be covariance stationary series, which is usually called static time series. As most economic time series do, if a time series we interested in are non-static, its first order difference should be static.

As the spectral representation theorem says, any covariance process can be represented as follows.

\[ y_t = a_0 + \sum_{j=1}^{M} [a_j \cos(\omega_j t) + \delta_j \sin(\omega_j t)] \]
are both random variables with zero mean.

1) Unique Spectrum Analysis

Choose any points from time series, the following are the calculation methods of the spectrum s at the point j.

First get the autocorrelation coefficient.

\[ C_k = \frac{1}{n-k} \sum_{t=1}^{n-k} \left( \frac{x_t - \bar{x}}{s} \right) \left( \frac{x_{t+k} - \bar{x}}{s} \right) \]

\( n \) is sample size, \( m \) is the number of points and \( \bar{x} \) is the mean of \( x \).

\[ s = \sqrt{\frac{1}{n} \sum_{t=1}^{n} (x_t - \bar{x})^2} \quad t=1, 2, \ldots, n; \quad k=0, 1, \ldots, m. \]

Then, calculate spectrum.

\[ \begin{aligned} 
    s_j &= (B_j / m) C_0 + 2 \sum_{k=1}^{m} C_k \lambda_k \cos(\pi j k / m) \\
    B_j &= \begin{cases} 
        1 & (j = 1, 2, \ldots, m-1) \\
        \frac{1}{2} & (j = 0, m) 
    \end{cases} 
\end{aligned} \]

2) Cross Spectrum Analysis

Given time series \( x \) and \( y \), the autocovariance coefficients at lag \( k \), cospectrum and quadrature spectrum (Q) of the cross spectrum are represented as follows.

\[ C_{xy}(k) = \left( \frac{1}{n-k} \right) \sum_{t=1}^{n-k} \left( \frac{x_t - \bar{x}}{s_x} \right) \left( \frac{y_{t+k} - \bar{y}}{s_y} \right) \]

\[ C_{yx}(k) = \left( \frac{1}{n-k} \right) \sum_{t=1}^{n-k} \left( \frac{x_{t+k} - \bar{x}}{s_x} \right) \left( \frac{y_t - \bar{y}}{s_y} \right) \]

Amplitude (A), phase (H), coagulation spectrum (O), lag (L) of the cross spectrum are as follows.

\[ \begin{aligned} 
    A_j &= P_j^2 + Q_j^2 \left( \frac{1}{2} \right) \\
    H_j &= \tan^{-1}(-Q_j/P_j) \\
    O &= A/(s_x s_y) \\
    L(T_j) &= P_j \left( \frac{T_j}{2\pi} \right) 
\end{aligned} \]

\( T_j \) is cycle length of the cross spectrum.

In the paper of the analysis of correlation relationship among international futures markets, Ye H. [11] applied the above methods and chose copper futures data as the study object. He chose the copper futures price data from LME and SHFE with a contract period of 3 months in 2008 and finally got 237 pairs of data by excluding the mismatched data.

First by correlation analysis, it was obviously that price fluctuation of copper futures in SHFE and LME has a high correlation relationship. After ADF and co-integrated tests, he found that there was co-integrated relationship between the two series. Then Granger causality test indicated that LME plays a guiding role in copper futures price between LME and SHFE.

The results showed that cycle for the copper futures price fluctuation of SHFE and LME with 9.82 days had the highest correlation. Corresponding frequency is 158, coherency coefficient is 0.802 and phase-different is -0.65.

Leading time of this cycle is -1.0164, i.e. the price fluctuation of LME comes 1.1064 days before SHFE. The result that the copper futures price fluctuation of LME plays a guiding role in the two markets is consistent with the outcome of Granger causality test. Copper price fluctuation of LME can be used to forecast the fluctuation of SHFE the next day.

Using this sequence of methods can help us to find out the correlation relationship between any two series. If there exists co-integration relationship, cross spectrum analysis will figure out the parameters for prediction model. By this model, we can get the prediction for the series we want.

D. Arbitrage

Compared with stock market, futures market is a better place for arbitrage. There are four main characteristics.

Firstly, also the biggest advantage, futures market supports two-way trade, that is to say, there is short-mechanism in the futures market. If the price is predicted to be rising, you can buy futures and sell it after the price rising. In contrary, if you believe that the price will decline, then sell futures immediately and buy them at a lower price. Therefore, speculators can find the arbitrage opportunities under both two price trends.

Secondly, there is low transaction cost in futures market. Commission charge is the only charge in futures market because stamp tax imposed in stock market does not exist here. The total one-side commission charge is less than 1‰.

Thirdly, using of leverage is the fascination of futures investment. It is not necessary to pay the whole money but a little margin. In China, 5‰ margin will be paid sometimes. In this case, market quotation is expanded by more than ten times.

Lastly, “T+0” transaction model is used in futures transaction. High speed trading is convenient because a futures contract can be bought and sold in one day.

It is very possible for the speculators to realize arbitrage under these favorable conditions, along with the correlation among markets. Since we find the correlation between two futures markets, futures price fluctuation of the guided market can be predicted by using the price data of the guiding market.
Here, as the example above, we can predict copper futures price data fluctuation in SHFE with the data price in LME the day before. For example, if the predicted price trend of copper futures in SHFE is rising, we can buy some copper futures contracts and hold them until the peak comes.

Hence, with price fluctuation data of the guiding market a unit of leading time before and the characteristic of futures market, arbitrage in the guide market is feasible.

IV. ANALYSIS SYSTEM

A. Framework and Process

Analysis system is used to analyze the correlation relationship between any two price fluctuation time series and forecast the price trend of the guided market, based on the correlation model mentioned above.

Once accepted the price data that user chooses, our system start data processing. First step is data validity test and the second is spectrum analysis, which will help us to find out the estimations of the prediction model.

Based on the model it produced and the history price data of the guiding market, the price trend of guided market can be predicted. The framework of the system is as Fig.1.

B. Modules

This system is composed of seven modules: data importing, sample selection, database management, knowledge base management, model base management, time series analysis and suggestion output. The whole structure is showed in Fig. 3.

Data importing imports the external price data into analysis system and converts them to the proper format. Sample selection checks the time series and excludes mismatched data. Database management deals with data inserting, updating, deleting, querying and backing up. Knowledge base management and model base management are used to manage the knowledge and model constructing, storing and updating respectively. Time series analysis uses econometrics methods to find the correlation between time series and analyzes the deviation between the prediction and actual results. Suggestion output gives analysis results and investment suggestion to the user.
By this system, user can scientifically know the correlation relationship between any related futures markets, get the price trend prediction of the guided market that he/she concerns, and then make the investment decisions to gain arbitrage.

V. COMPUTING TECHNOLOGY AND SYSTEM IMPLEMENTATION

A. Econometrics in R

As an implementation of the object-oriented mathematical programming language S, R is free software developed by statisticians around the world. It is very similar to S+, the most popular package used by econometricians.

R environment are composed of a series tools for data processing, computing and graphic display. Comparing to other similar software, it has a lot of characteristics.

First, there is effective mechanism for data processing and storage in R environment. Second, complete array and matrix operators make calculation for matrix become simple. Third, integrated data analysis tools provide an easy way to do economic analysis. Fourth, many kinds of graphic tools may directly deal with and display the data user concerns. The last but important one is that R is not just a software that does regressions and tests, but a modern mathematical programming language. When researchers develop new algorithms, the corresponding libraries will be posted in the R website. Because of the flexibility of R, it is easy to extend and always at the forefront of statistical knowledge. Moreover, code written for R can be run on any computational platforms such as .net, J2EE etc.

R is a wonderful tool for interactive data analysis. Because of the characteristics above, development cycle with R is usually short.

In our system, we can use R to do the time series analysis work on the .net platform. Using S to program at client and conveying it to software R for computing, at last the result will be return to C#. R(D)COM interface which can be used for C# to call R, enable one to use R as a computational engine that also can render graphics.

Fig. 4 shows the interaction between .net and R.

1) Libraries

First of all, we should construct the R(D)OM by installing package rsproxy. For using econometrics analysis we need, packages urca, uroot and tseries are necessary. RODBC is package for communicating with DBMSs.

Following COM references should be added to the C#.net project, as well as the namespaces to the class.

```csharp
using RServerManager;
using StatConnectorCommonLib;
using STATCONNECTORSRVLib;
```

After initialize the connector class of R, use EvaluateNoReturn() to casting and extracting data from the R environment. As we want to do economic analysis and database operation, the corresponding libraries should be added.

To add libraries, use method as following.

```csharp
StatConnector sc = new StatConnectorClass();
sc.Init("R");
sc.EvaluateNoReturn("library(urca)");
sc.EvaluateNoReturn("library(RODBC)");
```

2) Data Management

Data from text files, other statistical software and database are all possible to be imported into R environment.

To read the data from text file and database, use the following methods.

```csharp
//text files
sc.EvaluateNoReturn("m1=read.table(file.choose())");
sc.EvaluateNoReturn("library(RODBC)");
//database w/r
sc.EvaluateNoReturn("channel <- odbcConnect(\"d\", uid=\"ripley\")");
sc.EvaluateNoReturn("sqlFetch(channel, \"t\", rownames = \"state\")");
sc.EvaluateNoReturn("data(t)");
string r = "sqlSave(channel, t, rownames = \"s\", addPK = TRUE)"");
sc.EvaluateNoReturn(r);
```

3) Econometrics Analysis

The objective for using R in our system is to complete the data processing. To build prediction model, ADF test, co-integrated test, Granger causality test, cross spectrum analysis and deviation analysis is necessary.
It is very convenient to use R for econometrics analyzing. Only several statements are needed to complete a certain kind of statistical calculation. The following codes are used to do ADF test with lag 1.

```r
string rcode = "adf.out = unitroot(uscn.spot, trend="c", + statistic="t", method="adf", lags=1")" sc.EvaluateNoReturn(rcode);
```

As other examples, the major codes are "coint.rc = coint(uscn.ts,trend="c",lags=4)" is used for co-integration test and "spectrum( SOI, span = sqrt(2 * length(SOI)) )" is used for spectrum analysis.

4) Graphics

Graphical Tool is a very important component of R environment. We can use the tool to display all kinds of statistical graphic and create personalization graphics.

Interactions and batch processing of graphics is supported in R. Advanced drawing command will produce a new canvas, including coordinate axes, labels and title, etc.

The most commonly used graphics function is plot(). For example, if x and y are vectors, plot(xy) will produce a scatter.

By using graphics tool, almost all kinds of statistics graphics can be displayed.

Interactions between .net and R provide us a more convenient way to do economic analysis. The time series analysis module uses R to realize functions of price data processing, such as ADF test, co-integrated test, Granger causality test and cross spectrum analysis. By the estimations calculated from the functions, the prediction model is established. After importing the history price data of the guiding market, price trend prediction of the guided market will be displayed in a chart.

B. Evaluation

We have evaluated this system by a series of tests. Here we set the copper futures of SHFE and LME as an example to evaluate the system because some researches implied a relationship of futures price fluctuation between them. The copper futures price data of SHFE and LME from 1 April to 20 May in the year 2008 are chosen to predict the price trend of 21 May in the guided market.

Fig. 5 shows the prediction of copper price trend in guided market LME on 21 May, 2008. We can see from the figure that the leading time turns out to be 1 hour. That is to say, in this example, SHFE is the guiding market and the copper futures price fluctuation of LME may be the similar as the fluctuation of SHFE an hour before.

To test the effectiveness of the prediction, we compare the observed data and the prediction data as Fig. 6. Comparing the position of scatters and the points on the line, we found that they are very close to the correspondent points. The prediction of price trend and peak position is essentially consistent with the observed data.

The above tests give the support that the prediction model can correctly present the correlation relationship between the markets we chose. One can use this model to forecast the price fluctuation in guided market. It is instructive in investing in the capricious futures markets.

By simple operation, this analysis system has accurately found the correlation between two price fluctuations data and well predicted the futures price trend in the guided market.
VI. DISCUSSION AND APPLICATION

Our research provides an analysis system of futures price fluctuation by combining information technology with prediction model which is based on economic analysis such as validity test, spectrum analysis and deviation analysis. It offers a new thinking to quantitative study in this field. Meanwhile, the introduction of R into programming platform makes complicated statistics calculation in information system more convenient.

The method for predicting the price trend here is not only used to forecast the price fluctuation of the futures mentioned in this paper but also other futures and provides theoretical support to other price fluctuation researches. Moreover, the interaction between R environment and programming platform extends the ways of economic analysis in information system. By applying this method, prediction model and analysis system may also be brought into other fields.

For investors, the analysis system provides an accessible tool to analyze the correlation between futures markets and make the investment decision scientifically by being aware of the prediction of price trend.

Introducing the analysis system into futures markets can also instruct the domestic agricultural and industrial production plan. As a result, a country may reduce risks, meet the need of market, gain economic benefits and form a mature futures market.

VII. CONCLUSION

Economic globalization makes international futures markets become integrated. In this environment, even slight fluctuation of price seems to be sensitive and price fluctuation of one market may cause the fluctuation of another market in a faraway country.

During to the sensitivity of price fluctuation, there are some high correlation relationships among markets. Guiding markets may play an important role in the futures price of the guided market. Thus, the aim of our research is to develop some scientific methods aided by an information system to find out the related markets, distinguish the guiding market from the guided market, figure out the leading date and other coefficients of prediction model and finally use the price data from guiding market to forecast the price fluctuation of the guided market.

In this paper, we present a series of methods to analyze the correlation relationship between two related futures markets and describe an analysis system based on the interaction between .net platform and R environment. The system offers a new to do statistical calculations and economic analysis, which makes the prediction of price trend so simple and convenient.

However, mechanism of futures price fluctuation prediction and arbitrage is still at the starting stage. It can be improved from many respects.

Firstly, more analysis methods can be introduced into the price fluctuation study, during to the imperfect of spectrum analysis used in our research.

Secondly, since there are various futures in international futures market, a lot of futures can be studied in further research besides copper futures mentioned in our study. Moreover, the achievement of study may apply to other fields such as spot market to learn more about the internal relations among global market.

Thirdly, in order to get more accurate prediction model, more precision of price data such as hourly price data may be better than the daily price data used here.

Lastly, more study on the analysis system for arbitrage program or even transaction robots are also further research work.

ACKNOWLEDGMENT

This research is supported by Shanghai Leading Academic Discipline Project (No.B210).

REFERENCES

Weihui Dai received his B.S. degree in Automation Engineering in 1987, his Master degree in Automobile Electronics in 1992, and his Ph.D. in Biomedical Engineering in 1996, all from Zhejiang University, China. He worked as a post-doctor at School of Management, Fudan University from 1997 to 1999, a visiting scholar at Sloan School of Management, M.I.T from 2000 to 2001, and a visiting professor at Chonnam National University, Korea from 2001 to 2002. He is currently an Associate Professor at the Department of Information Management and Information Systems, School of Management, Fudan University, China.

Dr. Dai has published more than 120 papers in Software Engineering, Information Management and Information Systems, Complex Adaptive System and Socioeconomic Ecology, Digital Arts and Creative Industry, etc. Dr. Dai became a member of IEEE in 2003, a senior member of China Computer Society in 2004, and a senior member of China Society of Technology Economics in 2004.

Xingyun Dai received her BS degree in Computer Science and Technology in 2008 from East China Normal University, China. She is current a master student at School of Management, Fudan University, China.

Hua Ye received his BS degree in Information Management and Information Systems in 2009 from School of Management, Fudan University, China.

Jie Liu received his B.S. degree in Industrial Automation in 1987, his Master degree in Industrial Automation in 1990, and his Ph.D. in Management Science and Engineering in 1995, all from Tongji University, China. He is currently a Professor at the Department of Information Management and Information Systems, School of Management, Fudan University, China.