

# ARIMA Analysis of the volatility of the Soybean Futures Index of Dalian Commodity Exchange

Jincheng Huang<sup>1, a</sup>, Yan Li<sup>2, b</sup>

<sup>1</sup> Statistics Department, Faculty of Science National University of Singapore, 118426, Singapore

<sup>2</sup> Corresponding author, Management College Dalian Maritime University, Dalian, 116026, China

<sup>a</sup> 2466285149@qq.com

<sup>b</sup> lillyanfire@163.com

**Keywords:** Soybean Futures Index; ARIMA; Volatility; Dalian Commodity Exchange

**Abstract.** The futures market is an important part of the economy. It has the functions of price discovery, risk aversion and asset allocation. This article is a summary of related studies of soybean futures index using No. 1 Soybean Index as the subject of research. 2813 pieces of raw data of Nanhua Soybean Index were selected, and the trend of Nanhua Soybean Index is fitted using ARIMA model. The results showed that the ARIMA model is suitable for extracting the information from the raw data, and thus it is highly essential for the analyses of future indexes. It is an important tool for the in-depth studies of financial derivatives.

## 1. Introduction

In the nineties decades of the 20<sup>th</sup> century, China's Financial Futures Market acquired considerable development from Foreign Exchange Futures trading, Treasury Bond Futures trading and etc<sup>[1]</sup>. Although the development of China's Futures Market then had to face a pause for many reasons, in April 2010, Stock Index Futures revived in China Financial Futures Exchange, and in September 2013, Treasury Bond Futures regain its vitality after 18 years' absent. See Table 1.

China Financial Futures Exchange was established in Shanghai, which marks the milestone of the development of China's financial futures market. The pioneering Stock Index Future of China Financial Futures was officially listed on 16 April 2010, which fulfills the blank of China's Stock Index Futures trading market. Meanwhile, the re-appearance of Treasury Bond Futures also gave vitality to the financial futures market.

Stock Index futures has two characteristics, namely stock factor and futures factor. It is one of financial derivatives, and thus it is a product based on foundation products or basic variables. It is a corresponding concept to foundation financial product. It's advanced financial product whose price is largely determined by the price change of foundation financial product.

Table 1 China's Financial Futures History

Category	Duration	Summary	Reason of the break
Treasury Bond Futures	1992.12-1995.05; 2013.09 until now	Developed so rapidly in 1990s that it became a fashion; it was shutdown half-way	Several events led by "327 Treasury Bond Futures" became disturbance of the market.
Foreign Exchange Futures	1992.06-1996.01	The trading market was depressed	Short of amount of market trading.
Stock Index Futures	1993.03-1993.09; 2010.04 until now	The function of risk averse could hardly be utilized back in 1993.	In the 1990s, large amount of traders suppressed the index

Stock Index Futures were well developed in Occident, for instance, the Dow Jones and S&P500 Index Futures. International Stock Index Futures has some clear traits, which can be seen from the developing trends of the international market, like continuously updating new products,

simultaneously undergoing different electronic trading innovation. China's Stock Index Market is an emerging market, so it needs more studies on its volatility and its tendency.

With the rapid development of the futures market of agricultural commodity in China, the increasing number of types of futures trading and the growing trading volume<sup>[2]</sup>, the price discovery and risk aversion functions of futures market is getting improved, and hence its ability to stabilize market prices is showing a demonstrable effect.

Since the commodity futures index was first introduced half a century ago, it has played an important role in both the commodity market and the macroeconomics<sup>[3]</sup>. Soybean futures is listed in the early stage of the agricultural product futures market in china, and it is the variety with the highest trading volume and the strongest pricing ability. Due to the relatively mature market system and huge trading volume of soybean futures, it is sensitive to market information. Under the 'New Normal' state of China's economy environment, the volatility of the international bulk agricultural commodity prices has become significantly higher under the combined effect of international speculative capital, natural disasters, financial crises and the reduced stocks supply.

The study of futures price is the core of the study of futures, hence, an emphasis has been placed on the study of the price volatility, and there are numerous valuable research results. There are non-stationarity and randomness in futures market, so using only auto regression model or moving average model might not be comprehensive<sup>[4]</sup>. However, the differencing process in ARIMA model is able to filter out the non-stationarity, and ARIMA model mixes the advantages of the two models. Therefore, ARIMA model fits futures price better. George E.P.Box and Gwilym M.Jenkins published their pioneering book about time series (1970) and first introduced ARIMA model (also called Box-Jenkins model). The model has considered the stationarity of the time series, AR process and MA process. It uses p, d, q three parameters to amend the model. He G.G, Li Y. And Ma T.F. applied ARIMA model to study Short-Term Traffic Flow<sup>[5]</sup>; Lei K.W., Chen Y. used both ARIMA and BP nervous system model to study the number of inbound tourist in China<sup>[6]</sup>.

## 2. Construction of Arima Model

ARIMA model (Autoregressive Integrated Moving Average Model) is used throughout this study, where AR model stands for the autoregressive model with autoregressive order p; MA is the Moving Average Model with moving average order q; d is the degree of differencing which has to be done to obtain a stationary time series<sup>[7]</sup>.

ARIMA model is made up by following parts:

1. Differencing: If we start off with non-stationary raw data, several times of differencing have to be done to the time series to obtain a stationary time series.

2. AR Process: If an AR model with order p fits the time series, we have

$$x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + \phi_3 x_{t-3} \dots + \phi_p x_{t-p} + \omega_t \quad (1)$$

Where  $x_t$  stands for the value of the time series at time t,  $\phi$  s are p constants where  $\phi_p$  is not 0. The autocorrelation function cuts off after lag p.  $\omega_t$  is a time series which consists of collection of uncorrelated random variables.  $E(\omega_t) = 0$

3. Moving Average Process: If an MA model with order q fits the time series, we have

$$x_t = \omega_t + \theta_1 \omega_{t-1} + \theta_2 \omega_{t-2} + \theta_3 \omega_{t-3} \dots \theta_q \omega_{t-q} \quad (2)$$

Partial autocorrelation function cuts off after lag q.  $\theta$  s are q constants where  $\theta_q$  is not zero.

Finally, combining all the information, we can obtain an ARIMA model with order (p, d, q). In this model, the direct relationship between value and time is unseen, but expressed by AR process and

MA process which state this relationship indirectly<sup>[8]</sup>. After d times of differencing, we have the following fitting model:

$$x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + \phi_3 x_{t-3} + \dots + \phi_p x_{t-p} + \omega_t + \theta_1 \omega_{t-1} + \theta_2 \omega_{t-2} + \dots + \theta_q \omega_{t-q} \quad (3)$$

Different time series may result in different values of p, d, q. Different values of p, d and q will give different expressions of the model. The equation will have p number of AR coefficients,  $\phi$  and q number of MA coefficients,  $\theta$ . In application, values of p, d and q are small. Normally, d is not greater than 2 and p, q are not greater than 4.

To test stationarity and construct the fitting model, the above steps must be followed in application of ARIMA model<sup>[9][10]</sup>. Several methods may be used to obtain the coefficients: method of using ACF and PACF graph, method of exclusion, method of constructing and combining, method of enumeration, etc.

### 3. Positive Analysis

Nanhua index is an important index which Nanhua Futures Co., Ltd. uses to evaluate the commodity futures. In this research, 2813 pieces of raw data, between 01 June 2004 and 18 December 2015, have been selected from the Nanhua Soybean Index.

In R language, the daily data of the time series graph of soybean index can be obtained (Fig. 1). In general, the raw data shows a high volatility. High volatility can be observed between the  $t=1$  and  $t=1100$  interval, and no clear trend is seen. The interval between  $t=1100$  and  $t=2500$  shows a clear increasing trend. While a decreasing trend is observed after  $t=2500$ . The irregular volatility implies randomness.

The sample's non-stationarity can be shown by the Unit Root Test; therefore, a differencing process is needed. After 1st order differencing, the descriptive statistic is as follows: 1st order differencing; minimum -86.42; maximum 77.65; mean 0.1982; medium 0.37; number of observations 2812; variance 250.62 (Fig. 2). Using the adfTest command in R language to conduct the Unit Root Test, the p-value obtained is much less than 0.01, therefore,  $H_0$ , the time series is non-stationary, is rejected, thus this time series is thought to be stationary. Therefore, the 1st order differencing of the raw data has resulted in the stationary time series needed, the value of d is confirmed to be 1.

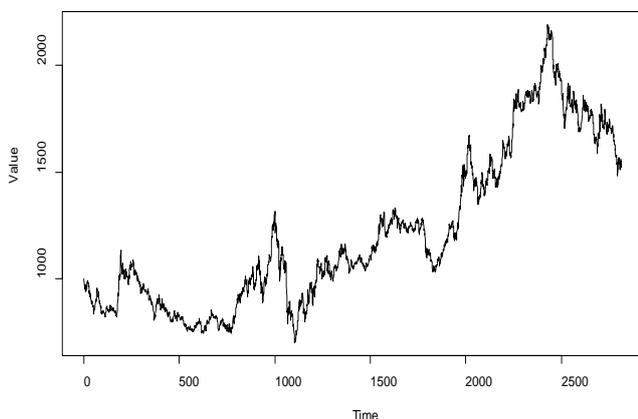


Fig. 1. Time Series Graph of Nanhua Soybean Index (1 June 2004,  $t=1$ , to 18 December 2015,  $t=2813$ )

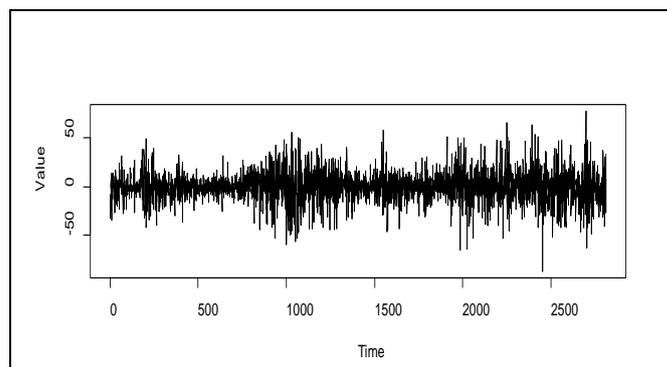


Fig. 2. 1st order differencing of Nanhua Soybean Index

The autocorrelation coefficient approximates to 0 after lag 1. The values of the autocorrelation coefficients from lag 1 to lag 35 are all in the 95% confidence interval. (Fig. 3)

The partial autocorrelation coefficients cut off after lag 2. Although the value of partial autocorrelation coefficient at l=9 exceeds the upper limit of the 95% confidence interval, only 1 of the 33 coefficients has exceeded the upper limit. The frequency of occurrences of extreme values is still less than 5%. Hence,  $H_0$ , the coefficients are 0, is not necessarily rejected. (Fig. 4)

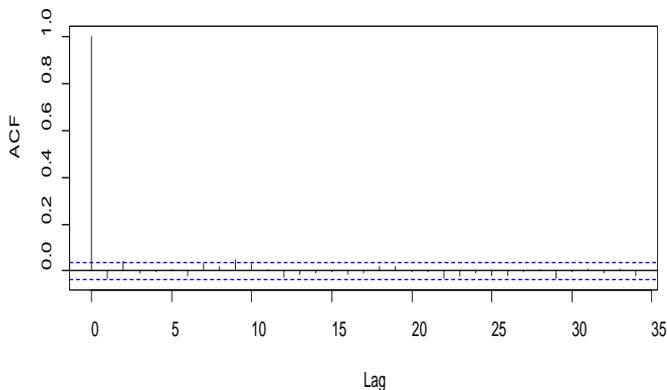


Fig. 3. Graph of the autocorrelation coefficients of Nanhua Soybean Index from lag 1 to lag 35

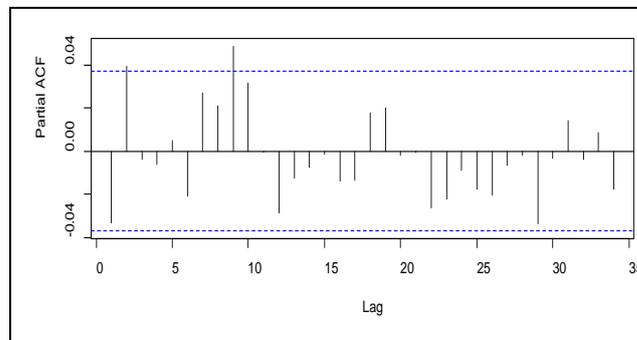


Fig. 4. Graph of the partial autocorrelation coefficients of Nanhua Soybean Index from lag 1 to lag 35

It can be seen from the ACF graph and PACF graph of 1st order differencing series that the PACF cuts off after 1 lag and ACF tails off. According to the BIC criteria, it can be confirmed that an ARIMA (2,1,0) model should be constructed to analyze and forecast the soybean index. The calculation using R shows that  $\phi_1 = -0.0317$  and  $\phi_2 = 0.0397$ . Fitting model:

$$x_t = -0.0317x_{t-1} + 0.0397x_{t-2} + \omega_t \quad (4)$$

Because the value of  $\phi_1$  is smaller than 0, the consecutive values of the function appear slightly negatively related. The predictions of the futures indexes between  $t = 2814$  and  $t = 2832$  interval using this model are as follows (Table 2) (Lo 80 and Hi 80 stand for the lower bound and upper bound of confidence interval of 80% confidence level, Lo 95 and Hi 95 stand for the lower bound and upper bound of confidence interval of 95% confidence level):

Table 2 Forecast of Nanhua Soybean Index(t=2814 to t= 2823)

Point	Forecast	Lo80	Hi80	Lo 95	Hi 95
2814	1555.30	1535.0	1575.6	1524.3	1586.3
2815	1556.70	1528.5	1584.9	1513.6	1599.8
2816	1556.58	1521.7	1591.4	1503.3	1609.8
2817	1556.63	1516.3	1597.0	1494.9	1618.4
2818	1556.63	1511.4	1601.9	1487.5	1625.8
2819	1556.63	1507.0	1606.3	1480.7	1632.5
2820	1556.63	1502.9	1610.3	1474.6	1638.7
2821	1556.63	1499.2	1614.1	1468.8	1644.4
2822	1556.63	1495.7	1617.6	1463.4	1649.8
2823	1556.63	1492.4	1620.9	1458.3	1654.9

#### 4. Conclusion

This research uses ARIMA model to fit the trend of the Nanhua Soybean Index and take advantage of ARIMA's good precision on short-term forecast to predict the values of the time series from  $t = 2814$  to  $t = 2832$ . Method of ACF and PACF graphs is used to calculate the values of p and q. This method is easy to apply and it make easy understanding how p, q are selected. Researches indicate that

short-term forecast using ARIMA model gives precise results and narrows confidence intervals. If further forecast is to be made, precision may decrease quickly and the error may be enlarged. ARIMA model can be a good tool to abstract important information of the trend of futures price, and is helpful for deeper study of the financial derivatives market.

## 5. Acknowledgment

The paper is subsidized by Social Science Fund Projects in Liaoning Provincial (Grant No. L14AJY001); The Fundamental Research Funds for the Central Universities (Grant No. 3132016037).

## 6. References

- [1] Jiang Yuexiang, Gong Lei, Long Huaigang. Expiration Day Effects of CSI 300 Index Futures: Evidence from Index and Individual Stocks. *Journal of Zhejiang University (Humanities and Social Sciences)*. 2016(7):184-200.(In Chinese)
- [2] Wang Jinzhong, Liang He, Li Junlin. The fractal characters of the csi 300 index futures market research. *Economists*, 2015(12):60-67.(In Chinese)
- [3]The people's bank of China Zhengzhou central sub-branch research group, Study on macroeconomics warning function of futures index of China, *Financial Theory & Practice*,2008 (9): 3-9
- [4]Liu Chengyi. Markov chain short-term forecasting effectiveness based on the csi 300 index futures. *Financial and economic* .2016(10):71-73.(In Chinese)
- [5]He Guo-Guang, Li Yu, Ma Tao-Feng. Discussion on Short-Term Traffic Flow Forecasting Methods Based on Mathematical Models. *Systems Engineering-Theory & Practice* 2000, Vol 12:P51-56.(In Chinese)
- [6]Lei Ke-Wei,Chen Ying. Forecast of number of inbound tourist in China based on BP nervous system and ARIMA model, *Tourism Tribune* 2007 Vol 4-22:P20-25.(In Chinese)
- [7]Wang Xi-Tao. Study on the Application of ARIMA model in time - bargain forecast. *Microcomputer Information*2006-05-30. P139-140.(In Chinese)
- [8]Jonathan D.Cryer; Kung-Sik Chan. *Time Series Analysis with Applications in R* (2nd ed.). Springer. 2008
- [9]Li Zhan-Jiang, Zhang Hao, Sun Peng-Zhe etc. Study on CSI300 stock index futures prices based on ARIMA model. *Ludong University Journal( Natural Science Edition)*. 2012,29( 1) : P22—24.(In Chinese)
- [10]Zhang Wei, Zhang Yan-Qi, Yang Xu. Model of multiple seasonal ARIMA and its application to data in time series. *ACTA ACADEMIAE MEDICINAE MILITARIS TERTIAE*. 2002 Aug.Vol24-8 P955-957.(In Chinese)