Futures Arbitrage of Different Varieties and based on the Cointegration Which is under the Framework of Bayesian—In the Case of Soy Oil and Palm Oil

Bing-ting FAN

Shanghai University, 99 Shang Da Road, Baoshan District, Shanghai City, China
1973431340@qq.com

Keywords: Cointegration Model; Expected Return Maximization; The Bayesian Framework; Arbitrage Between Different Varieties.

Abstract. This paper mainly studies commodity futures arbitrage opportunities across the soy oil and palm oil. The traditional way is traders can carry into the market when the spread out of the reasonable scope. When the price back to normal levels, unwinding operation and complete an arbitrage process. However, spreads of soy oil and palm oil does not meet the stability, so we need to test cointegration model residual error after regression, combined with the maximization of expected return to determine the arbitrage opportunity. In traditional cointegration model, long-term equilibrium relationship is \( y_t = 3070.8 + 0.50x_t \). When the residual absolute value is greater than 61.125, arbitrageurs may take positions into the market, until the residual back to a reasonable range. However the traditional cointegration model without considering the estimation error, and estimated parameter has high sensitivity. So we will use the bayesian method to estimate parameters, we get the Bayesian-cointegration is \( y_t = 3071 + 0.54x_t \). The residual under this model is stable. Therefore, we can find arbitrage opportunities through the model. When the residual is more than 142.475 or less than -126.475 triggered buying conditions, and can gain maximum expected return.

Introduction

Zhengzhou grain wholesale market in 1990 formal introduction of futures trading mechanism, Help opens the door of the futures market in our country. Since 1993, the state council notice regulates the market and futures market gradually into the standard stage of development, China's futures market walked through the 23 years of wind and rain.

Our country as large agricultural products consumers, the analysis of agricultural products futures arbitrage opportunities across varieties has the practical significance, arbitrage make the market price is more reasonable, to enhance the stability and liquidity of the market. Soybean oil and palm oil are two important oil products, they have a strong alternative in the daily consumption and use. So they are in the futures market showed a strong correlation. Because the cointegration model without considering possible nonlinear relationship between the two commodity futures and it cannot consider asymmetric volatility near the error of the mean. So some scholars on the basis of traditional cointegration model introducing the threshold, and combining the fundamental analysis and technical analysis to the agricultural product futures market arbitrage opportunities to make judgment. This paper is to refer to previous experience, combined with bayesian framework to get stable cointegration model and introducing threshold judgment arbitrage opportunities.

Data

In order to avoid the risk, liquidity risk and delivery to ensure the efficiency of the model, this paper use of soybean futures and the main palm oil futures contracts data for empirical analysis. We will choose active contract settlement price of soy oil and palm oil commodity futures on May 4, 2015 to May 9, 2016. Correlation between commodity futures can make cross varieties arbitrage, and in this article, their correlation coefficient is 0.93, so the strong correlation between them. We need to judge the stationarity of soy oil and palm oil prices sequence before modeling, We can get
the results through the unit root test: P value of soy oil and palm oil are both 0.8, both are far greater than 0.05 and both are non-stationary sequence.

Mean-reverting process

On the market that decided by the power of supply and demand, The price of any asset is unlikely to present unilateral trend for a long time. The market price of the asset either above or below its intrinsic value will reply to its intrinsic value with large probability, this is the theory of mean reversion. It is an important theoretical basis across a variety of arbitrage.

Cointegration regression analysis

In an economic system, although various economic variables have their long-term variation rule due to different influence factors, but they maybe exist stable linear relationship, to show the non-stationary economic variables have a long-term and stable relationship. According to the result of unit root test determine soy oil and palm oil commodity futures contract data is non-stationary, so whether there is a linear combination between the two contracts is smooth, it is need to set up the cointegration regression to observe. $x_t$ means sequence data of palm oil , $y_t$ means sequence data of soy oil. Cointegration regression model can be set up as follows:

$$y_t = \alpha + \beta x_t + \varepsilon_t$$

Using OLS estimates can obtained the model $y_t = 3070.8 + 0.5x_t$. For cointegration model, residual stability determines whether there is a long-term equilibrium relationship between the two contracts. In this model, the residual error sequence through the stationary test, the p value is 0.0029, is far less than 0.05, it shows that spreads has strong ability of mean reversion.

Cointegration under the framework of bayesian

The classical cointegration model use sample estimates of parameters as if they were the true parameters, but its sensitivity is higher. So we can use the Bayesian framework to solve this sensitivity. The bayesian framework is a trade-off between the sample information and prior distribution, which is under the background of probability knowledge. Soy oil and palm oil commodity futures contract price there is a cointegration relationship as $y_t = \alpha + \beta x_t + \varepsilon_t$ . Parameter space is $\Theta = (\theta_1, \theta_2), \varepsilon_t \sim N(0, \tau^{-1}), \theta_1 = [\alpha, \beta]$. Assume that the prior distribution of $\theta_1 = [\alpha, \beta]$ is noninformative prior distributions, as $\pi (\theta_1) = 1$ . $\tau$ follows the Gamma distribution, as $\pi (\tau) \propto \tau^{\epsilon - 1}$. Each parameter is independent of each other. Joint prior distribution can be expressed in the following way: $\pi (\Theta) \propto \tau^{\epsilon}$.

When the parameters are known, the joint likelihood function can be written as :

$$L(\Theta, X, Y) \propto \frac{\tau}{\sqrt{2\pi}} \exp \{\frac{\tau}{2}(Y - \theta_1 X)'(Y - \theta_1 X)\}$$

the a posteriori probability of parameters is

$$\pi (\Theta | X, Y) \propto \tau^{\epsilon} \frac{\tau}{\sqrt{2\pi}} \exp \{\frac{\tau}{2}(\hat{\theta}_1 - \hat{\theta}_1)'X'X(\hat{\theta}_1 - \hat{\theta}_1)\}$$

And the marginal posterior distribution are given ,respectively, by

$$\pi (\alpha | \beta, \theta_2, X, Y) \propto \tau^{\epsilon} \frac{\tau}{\sqrt{2\pi}} \exp \{\frac{\tau}{2}(\hat{\theta}_1 - \hat{\theta}_1)'X'X(\hat{\theta}_1 - \hat{\theta}_1)\}$$

$$\pi (\beta | \alpha, \theta_2, X, Y) \propto \tau^{\epsilon} \frac{\tau}{\sqrt{2\pi}} \exp \{\frac{\tau}{2}(\hat{\theta}_1 - \hat{\theta}_1)'X'X(\hat{\theta}_1 - \hat{\theta}_1)\}$$
\[ \pi (\theta_2, \theta_1, X, Y) \propto \tau^2 e^{2\tau \frac{\tau}{\sqrt{2\pi}} \exp\left( \frac{\tau}{2} (\theta_1, \hat{\theta}_1)^T X^T X (\theta_1, \hat{\theta}_1) \right)} \]

where \( \hat{\theta}_1 = (X^T X)^{-1}X^TY \).

Due to the complexity of distribution form, we use Gibbs sampling for parameter estimation, the iteration steps as follow:

Step (1) Set the initial state \( \Theta_0 = (\alpha^0, \beta^0, \tau^0) \)

Step (2) Sampling from the posterior conditional distribution
\[ \pi (\alpha | \beta^0, \tau^0, X, Y) \propto \tau e^{\frac{\tau}{\sqrt{2\pi}} \exp\left( \frac{\tau}{2} (\theta_1, \hat{\theta}_1)^T X^T X (\theta_1, \hat{\theta}_1) \right)} \text{ and get } \alpha^1; \]

Step (3) Sampling from the posterior conditional distribution
\[ \pi (\beta | \alpha^1, \tau^0, X, Y) \propto \tau e^{\frac{\tau}{\sqrt{2\pi}} \exp\left( \frac{\tau}{2} (\theta_1, \hat{\theta}_1)^T X^T X (\theta_1, \hat{\theta}_1) \right)} \text{ and get } \beta^1; \]

Step (4) Sampling from the posterior conditional distribution
\[ \pi (\tau | \beta^1, \alpha^1, X, Y) \propto \tau e^{\frac{\tau}{\sqrt{2\pi}} \exp\left( \frac{\tau}{2} (\theta_1, \hat{\theta}_1)^T X^T X (\theta_1, \hat{\theta}_1) \right)} \text{ and get } \tau^1; \]

Step (2) - (4) completed the first round of the iteration, repeat step (2) - (4), until the markov chain present stationary, figure 1 is Gibbs sampling principle diagram, Model can be expressed as \( y_t = 3071 + 0.54x_t \).

Figure 1. Gibbs sampling principle diagram.

The Determination of Arbitrage Opportunities

How to determine the arbitrage interval, we need to analysis residual. When the spread value deviates from the value of the long-term equilibrium modestly, in a smaller range, no arbitrage space, at this point should be kept short positions. When the price deviating from the equilibrium value reaches a certain degree, the existence of arbitrage opportunities, can build arbitrage portfolio. We need to build optimization model to determine the arbitrage interval, to determine the threshold value of entry and exit the market.

Optimal threshold of traditional cointegration model

An arbitrage will make a profit \( R_y = \theta \bullet (P_{s,t} - P_{s,m}) + \theta \bullet \beta_y (P_{s,m} - P_{y,m}) = \theta \bullet (e_{s,t} - e_{s,m}) \) where, \( P_{s,t0} \) means the price of soy oil at time \( t, \theta \) means amount of soy futures, \( \beta_y \) means the coordination coefficient. In order to ensure a positive earnings, can be set \( R_y \geq s \bullet k \bullet \delta_y \), where \( \delta_y \) is the standard deviation of the residual. The selection of the \( k \) determines the profitability, If \( k \) is larger, might miss other profit opportunities; If small \( k \) value, arbitrage opportunities maybe more, but the

323
transaction cost will increase. We assume that the arbitrage expected revenue function is 
\( E(k\sigma) = \theta \cdot R(k\sigma) \cdot \phi^{-1}(k\sigma) \), by maximizing the expected revenue, the appropriate \( k \) value can be obtained.

Without considering the transaction costs, \( k = 0.75 \) can maximize the expected revenue, namely when the residual is more than 61.125 or less than -61.125, can get the biggest arbitrage profits.

**Optimal threshold of Cointegration which under the framework of bayesian**

Cointegration which under the framework of Bayesian consider the estimation error, so this model is superior to the traditional cointegration model.
Without considering the transaction costs, \( k = 1.65 \) can maximize the expected revenue, namely when the residual is more than 142.48 or less than -126.48, can get the biggest arbitrage profits.

**Summary**

Due to strong correlation which create arbitrage opportunities between soy oil and palm oil, However, the spreads between the two contracts is not stationary, so need to using cointegration model to establish the long-term equilibrium relationship between the two contracts, and get a steady residual which can be seen as a kind of trading futures contracts. Traditional cointegration model parameters are sensitive, so we considering parameters estimation in bayesian framework, cointegration under the bayesian framework on the basis of the original sample combined with the prior distribution of the parameters, to improve the accuracy of parameter estimation and sensitivity. Furthermore, the article through the maximum expected revenue obtained threshold of long positions and short positions, arbitrage through the model that cointegration which under the framework of bayesian can gain higher returns and lower risks.

**References**


