Universe
Scientific Publishing

Research of Arbitrage based on Corn and Corn Starch Futures Contracts

Biyu Zhu[1], Mengdi Hao[2], Zixuan Wu[3], Mandi Yang[4], Wei Song[5], Zhenhan Fu[3] (Co-First Authors)

1. The University of Glasgow, Glasgow G12 8QQ, United Kingdom.
2. The University of Chicago, Edward H. Levi Hall, Chicago 60637, United States of America.
3. University College London, London WC1E 6BT, United Kingdom.
4. Xi'an Jiaotong-Liverpool University, Suzhou 215000, China.
5. Imperial College London, London SW7 2AZ, United Kingdom.

Abstract: This paper attempts to use Error Correction Model (ECM) to arbitrage in the futures market. Corn and corn starch futures contracts are chosen since there is a high correlation between them. A passive investment method is used to compare with the result of ECM to examine whether it is effective in making benefits. For risk measure, this paper uses Extreme Value Theory (EVT) and Monte Carlo to calculate the VaR value using the ratio derived above for risk management. The results show that with both higher return and lower risk, the model built by ECM performed better than the benchmark investment method.

Keywords: Error Correction Model; Future Contract; Extreme Value Theory; Monte Carlo

Introduction

The error correction model (ECM) was established by Hcndry and Andcrson in 1977 to explain the short-term fluctuation and deviation between two non-stationary time series. Song (2018) used the arbitrage combination of corn futures and corn starch futures to find an arbitrage opportunity. The model of GARCH was used to calculate the conditional standard deviation of the time series dynamics and then calculated the sequence of spreads, mspread(t), between the two indexes to find the arbitrary opportunity. The GARCH model cannot be applied to find the accurate optimal ratio between two futures, thus the ECM was used by the author in this situation. The innovation of this paper is that it effectively combines the trend of the price difference and the return of the price difference to carry out the arbitrage model research.

Due to the transformation and upgrading of economic structure, the price volatility of stocks, bonds, and commodities is high, and investment risks are relatively large. In terms of this situation, Nikos et al [1](2006)used GARCH to set open/close position and stop loss point of futures by calculating dynamic condition standard deviation. In addition, the innovation of this paper is that it combined the GARCH model and the neural network model, which ensures it can constantly change arbitrage criteria to find arbitrage opportunities.

The method of VaR is the loss distribution quantile of an asset combination, which is highly related to the high-dimensional integrals. While the Monte Carlo approach is an essential tool for high-dimensional integrals. However, the convergence speed of the traditional Monte Carlo approach is very slow. In terms of this background, Teng [2](2017) organized research to prove the spherical Monte Carlo approach is more effective than the traditional Monte Carlo approach. The sample he chose followed the importance sampling estimator proposed by Glasserman, Heidelberger, and Shahabuddin. To prove the practicality of this method the research applied the VaR model with heavy-tailed distributions to compare the calculation time of variance between the spherical model and traditional model.

Some previous researchers have widely used the combination of the GARCH model and the extreme value theory (EVT) to predict the extreme risk of the futures portfolio. Chen and Yu[3] (2019) integrated the APARCH model and EVT, and this combination can give a more accurate prediction of VaR due to EVT directly focusing on the tail nature of the return. In addition, two categories of EVT are the traditional model and peaks-over-threshold (POT) model respectively, and the POT model is regarded as more useful which also be applied in this paper. Subsequently, to test the validity of the APARCH model based on EVT, the authors used the HSI future market as the sample and selected the data from January 1, 2006, to December 31, 2012, from the Wind Finance Terminal. As a result, it showed that the model is suitable for modeling return series of volatility clustering and leverage effect. Besides, it also showed that this model is very suitable for measuring extreme risks.

EMPIRAL ANALYSIS: PORTFOLIO CONSTRUCTION

1. Benchmark

Since two futures contracts are chosen in this paper, the passive investment method depicts a 1:1 ratio of investment. Daily returns are calculated as LN(Pt)-LN(Pt-1). With the daily price series, we get the return as 0.0796746%.

2. ECM Model

(1) Data Selection

We obtained these data from Dalian Commodity Futures Exchange, from 3rd January 2017 to 28th December 2018, totaling 485 observations for corn starch and corn futures contracts, respectively.

- Contract 1 $\{\alpha_1, \alpha_2, \cdots\cdots, \alpha_n\}$
- Where α denotes the price of contract 1,"1...n" denote dates. And we make the beginning price $\{\beta_1\}$ of second contract equal to α_n. That is:
- Contract 2 $\{\beta_1, \beta_2, \cdots\cdots, \beta_n\} \times \alpha_n/\beta_1$ and so on:
- Contract 3 $\{\gamma_1, \gamma_2, \cdots\cdots, \gamma_n\} \times \alpha_n\beta_n/(\alpha_1\beta_1)$
- Contract 4 $\{\delta_1, \delta_2, \cdots\cdots, \delta_n\} \times \frac{\alpha_n\beta_n\gamma_n}{\alpha_1\beta_1\gamma_1}$
- Contract 5 $\{\varepsilon_1, \varepsilon_2, \cdots\cdots, \varepsilon_n\} \times \frac{\alpha_n\beta_n\gamma_n\delta_n}{\alpha_1\beta_1\gamma_1\delta_1}$ ...till sixtieth contract

Figure 1

(2) Stationarity Tests

Prices data are logged firstly. Then, using R program to undertake ADF test:

```
Title:
 Augmented Dickey-Fuller Test

Test Results:
  PARAMETER:
    Lag Order: 0
  STATISTIC:
    Dickey-Fuller: -21.0572
  P VALUE:
    0.01

Description:
 Thu Nov 21 14:43:49 2019 by user: lenovo

Warning message:
In adfTest(lrcs, lags = 0, type = c("nc")) :
  p-value smaller than printed p-value
```

Figure 2 ADF Test of Corn Starch Futures Index Price Sequence

```
Title:
 Augmented Dickey-Fuller Test

Test Results:
  PARAMETER:
    Lag Order: 0
  STATISTIC:
    Dickey-Fuller: -22.6787
  P VALUE:
    0.01

Description:
 Thu Nov 21 14:43:50 2019 by user: lenovo

Warning message:
In adfTest(lrc, lags = 0, type = c("nc")) :
  p-value smaller than printed p-value
```

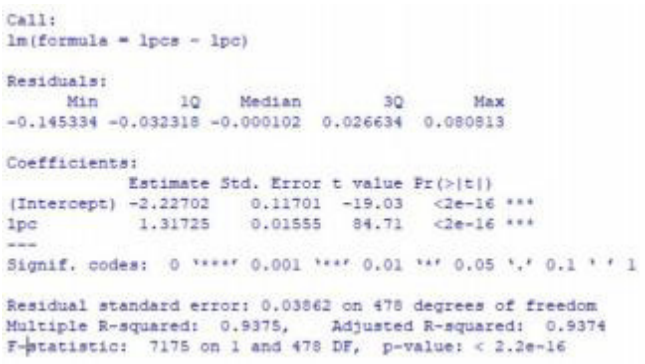
Figure 3 ADF Test of Corn Futures Index Price Sequence

As the ADF test result shows the null hypothesis (H0 : There are unit roots in the time series) is rejected.

(3) Co-integration Test

E-G two steps method is used to test the co-integration then:

a. Establishing the Co-integration Model

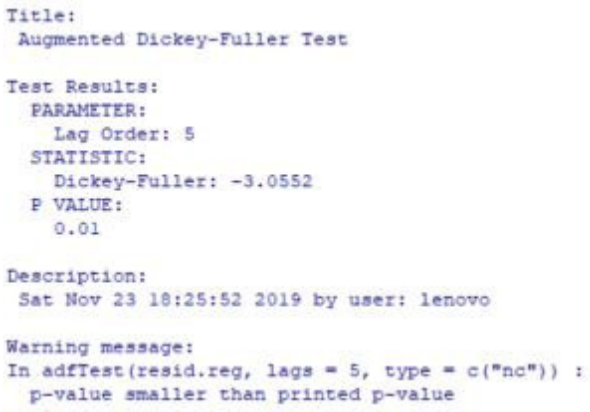


```
Call:
lm(formula = lpcs ~ lpc)

Residuals:
      Min        1Q    Median        3Q       Max
-0.145334 -0.032318 -0.000102  0.026634  0.080813

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -2.22702    0.11701  -19.03   <2e-16 ***
lpc          1.31725    0.01555   84.71   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.03862 on 478 degrees of freedom
Multiple R-squared:  0.9375,    Adjusted R-squared:  0.9374
F-statistic:  7175 on 1 and 478 DF,  p-value: < 2.2e-16
```

Figure 4 Co-integration Model of Corn Starch and Corn Futures Contracts

It is obvious that there is a linear relationship as LPCSt=- 2.22702+1.31725LPCt+μt, where LPCS and LPC denote logged price of corn starch and corn respectively. Then to test whether its residual is stationary. If stationary, there is a stable linear relationship between corn starch and corn.

b. Stationarity Test



```
Title:
 Augmented Dickey-Fuller Test

Test Results:
  PARAMETER:
    Lag Order: 5
  STATISTIC:
    Dickey-Fuller: -3.0552
  P VALUE:
    0.01

Description:
 Sat Nov 23 18:25:52 2019 by user: lenovo

Warning message:
In adfTest(resid.reg, lags = 5, type = c("nc")) :
  p-value smaller than printed p-value
```

Figure 5 ADF Test of Residual Series

According to the above ADF test result, the null hypothesis is rejected which shows the residual series is also stationary. This leads to the conclusion that there is co-integration between the price of corn starch and corn futures contracts.

c. Construction of ECM Model

Using the price of corn starch futures contracts as dependent variable, the price of corn futures contracts, the one-period lagged price of corn starch futures contracts and one-period lagged price of corn futures contracts as independent variables to construct the regression model:

```
Call:
lm(formula = lpcs ~ lpc + lpcl + lpcsl)

Residuals:
      Min        1Q    Median        3Q       Max
-0.049946 -0.003369 -0.000682  0.002879  0.081151

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -0.08341    0.03631  -2.297    0.022 *
lpc          0.51010    0.04574  11.153   <2e-16 ***
lpcl        -0.46666    0.04658 -10.018   <2e-16 ***
lpcsl        0.96839    0.01072  90.305   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.009043 on 476 degrees of freedom
Multiple R-squared:  0.9966,    Adjusted R-squared:  0.9966
F-statistic: 4.638e+04 on 3 and 476 DF,  p-value: < 2.2e-16
```

Figure 6 Error Correction Model

Since all the parameters in the model are significant, the equation below isderived:

Therefore, a short-term relationship between corn starch and corn futures contracts is:

To make sure all the information available is used in transaction, the ratio between corn starch and corn futures contract should be 1:1.37425.

Since spreadt=LPCSt-1.37425LPCt and to determine the trading signals, ARCH effect is tested to buildup GARCH model.

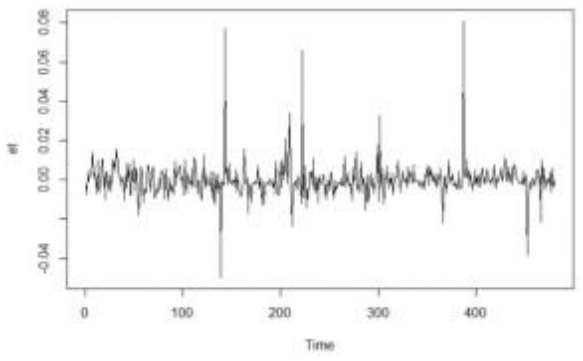


Figure 7 Plot of ECM Residuals

The above figure shows a high possibility of existing ARCH effect. Therefore, the next step is to test its residual of ARCH effect:

```
        Box-Ljung test

data:  at^2
X-squared = 8.9058, df = 4, p-value = 0.0635
```

Figure 8 ARCH Effect Test

According to the result of Ljung-Box test, the null hypothesis is rejected under significance level of 10% which shows there is ARCH effect in the residual series.

d. Establishing GARCH model

From GARCH model, we are able to derive dynamic conditional standard deviation of time series, based on which trading signals can be determined. If the price difference mspreadt exceeds its mean for a specified amount, it is said to have deviate the reasonable range, under which condition there is an arbitrage opportunity and an open signal is being sent to the investor. If it exceeds the mean for a huge amount, a close signal is sent to stop loss. When it returns into the reasonable range, a close signal is also sent to earn profits.

Using R program, the GARCH (1,1) model is derived as following:

```
Error Analysis:
        Estimate  Std. Error  t value Pr(>|t|)
mu     -1.312e-04   4.115e-04   -0.319   0.7499
ar1     1.000e+00   7.301e-02   13.697  < 2e-16 ***
ar2    -1.406e-02   9.127e-02   -0.154   0.8776
ar3    -5.690e-02   8.940e-02   -0.636   0.5245
ar4    -1.525e-01   7.529e-02   -2.025   0.0428 *
ar5     2.226e-01   5.281e-02    4.216 2.49e-05 ***
omega   3.158e-05   7.694e-06    4.104 4.06e-05 ***
alpha1  1.000e+00   2.072e-01    4.826 1.39e-06 ***
beta1   3.782e-01   4.582e-02    8.255 2.22e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Log Likelihood:
 1479.414    normalized:  3.082111
```

Figure 9 GARCH (1,1) Result

Thus, σ =-0.0001312+ ε−1 +0.3782 σ_{t-1}^2, where σ_t represents the conditional standard deviation of mspreadt at time t, σt−1 represents the one at time t- 1 and ε t represents the unconditional standard deviation of mspreadt at time t- 1. Next, it is time to set a series of criteria to give trading signals. According to Song[4] (2018), one unit of conditional standard deviation is appropriate to be the open signal and two units for close signal.

e. Back-test

From the test results, the annualized excess return can be calculated to be 1.06%, which shows that the arbitrage strategy is effective out of the sample period. In the back-test process, the initial funding is set to be RMB 1000,000. According to the hedging ratio derived from ECM, 137 units of corn futures contract are oppositely traded. The capital line displays a concussion uplink pattern and a low reverting range (which is all below 0.08). This shows that the return result by using the arbitrage strategy is effective and stable.

EMPIRAL ANALYSIS: RISK MANAGEMENT MODELS

1. Monte Carlo

We used “MASS” package of R program to perform Monte Carlo stochastic simulation of portfolio VaR. (Du,2018)[5].We simulated 5000 times with the help of “mvrnorm” multivariate variable simulation function and obtained 5000 randomly generated yield “logrt”, and saved the results as “dx” .

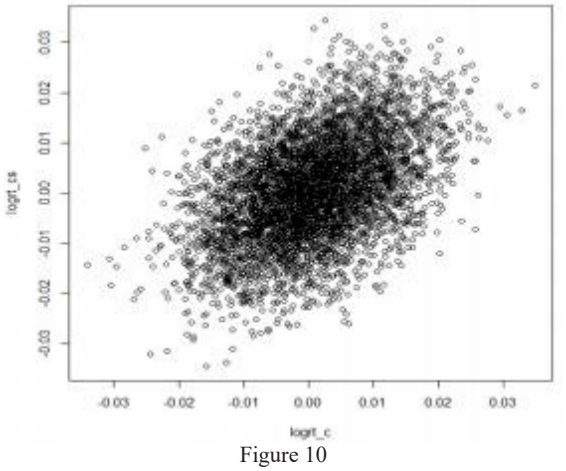


Figure 10

Then we combined the weights of the optimal portfolio calculated in the previous section with the simulation results and saved as “dp” .

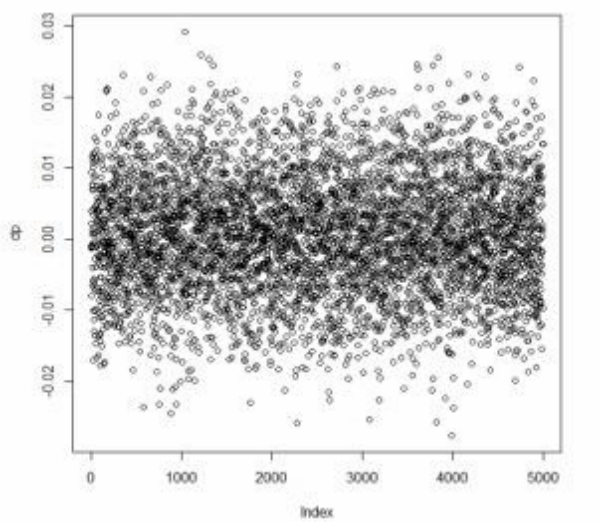


Figure 11

After that, we used the sort function to arrange the 5000 optimized results in descending order.

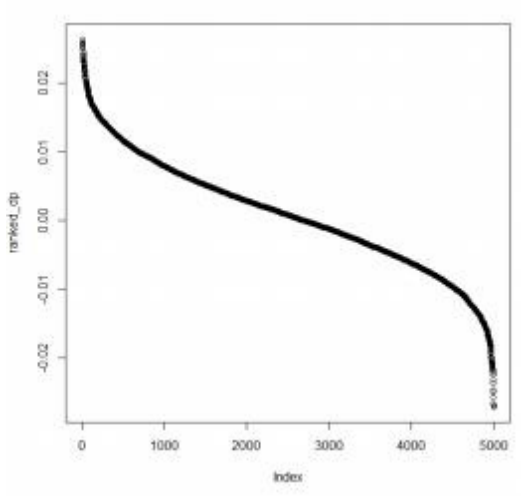


Figure 12

Next, we conducted similar process to the same historical data of corn future and corn starch future setting different confidence levels of 95% and 90%.

Below is the results summary of VaR and ES by Monto Carlo simulation:

Confidence Level	VaR (Benchmark)	ES(Benchmark)	VaR	ES
99%	1.920%	2.192%	1.945%	2.181%
95%	1.382%	1.717%	1.412%	1.773%
90%	1.430%	1.791%	1.405%	1.736%

Chart 1

In general, the traditional VaR calculation may underestimate the risk of return series, so we will next use the Extreme Value Theory to evaluate the risk.

2. Etreme Value Theory (EVT)

The hill plots of corn and cornstarch at 95% confidence interval are presented as below, respectively.

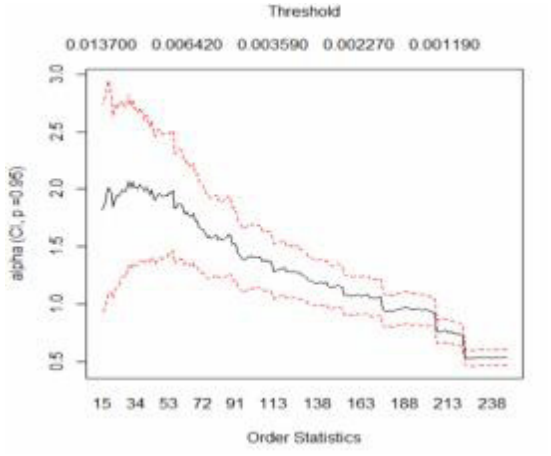


Figure 13

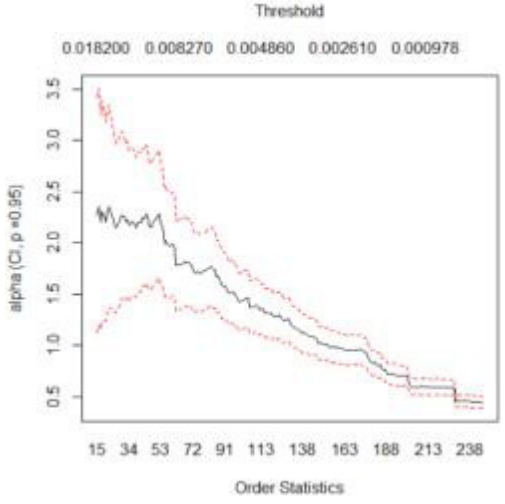


Figure 14

As the two figures shown, when the excess number is 53, the trend of log return tends to be chaos. Inputting the code of 'findthresh(l-rc,53)' in R program, the threshold is 0.00753772 and 0.01006098 for corn and cornstarch, respectively. The overall number of observations is 486 and the part of excess numbers accounts for the whole observations at 10.905%, which is reasonable in analysis. The Q-Q plot of the two datasets is quasi-concave, which indicates that the heavy tail exists(Voropaev, M., 2009.)[6].

1 Corn

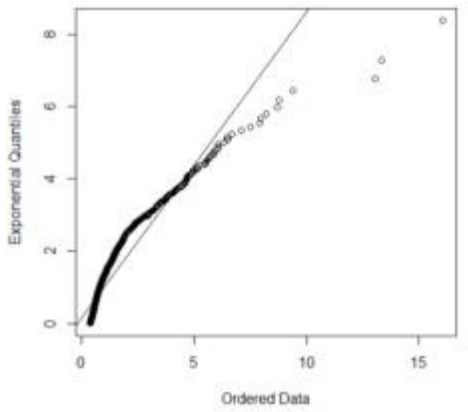


Figure 15

2 Cornstarch

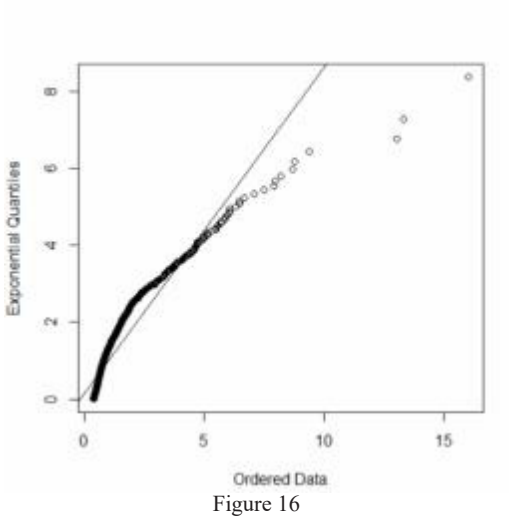


Figure 16

The empirical distribution plots are approximately straight. Therefore, these two groups of data indicate Pareto tail behavior.

3 Corn emplot

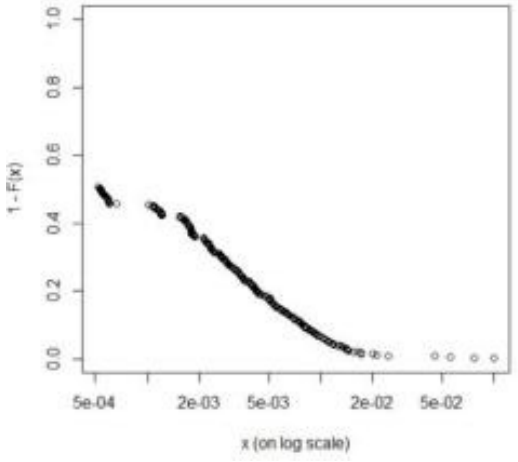


Figure 17

4 Cornstarch emplot

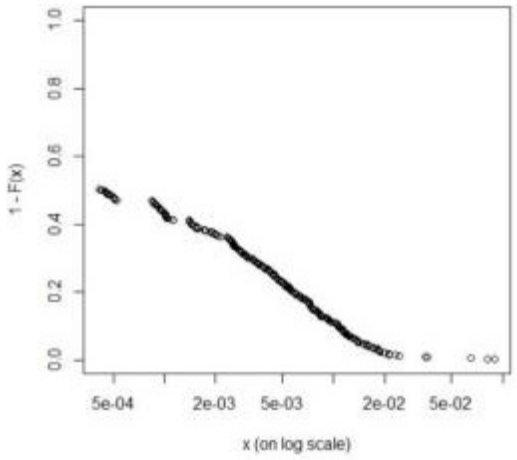


Figure 18

Combining the thresholds, the parameters of GPD can be calculated as the list showed.

Contract	Threshold	$\bar{\varepsilon}$	$\tilde{\beta}$
Corn	0.00753772	0.59438368	0.00361760
Cornstarch	0.01006098	0.48411930	0.00468537

Chart 2

Kolmogorov-Smirnov test is a test to compare a frequency distribution with a theoretical distribution or two observed distributions. Here is used to examine the degree of tail fitting. For the two contracts, the result of P-value is larger than 0.05 so that the null hypothesis of the tail fitting distribution of the sequence obeying uniform distribution cannot be rejected.Then the VaR based on the EVT can be calculated using the threshold and estimated parameters in GPD in R program. The result of risk measure is as listed.

Corn

Confidence interval	VaR	ES (CVaR)
99%	0.02663454	0.06353747
95%	0.01112647	0.02530413
90%	0.007859471	0.01724972

Chart 3

Cornstarch

Confidence interval	VaR	ES (CVaR)
99%	0.03115323	0.06002916
95%	0.01450003	0.02774805
90%	0.01047569	0.01994715

Chart 4

Weight of corn and cornstarch in the set portfolio is 57.88% and 42.12%,respectively. And thefinal calculation of VaR is according to the corresponding weight of the two futures.

Confidence interval	VaR	VaR(Benchmark)	CVaR	CVaR(Benchmark)
99%	2.854%	2.889%	6.21%	6.18%
95%	1.255%	1.281%	2.63%	2.65%
90%	0.896%	0.917%	1.84%	1.86%

Chart 5

Portfolio construction model and Risk models performance evaluation

As chart 1 and 5 show, the risk of the constructed portfolio is lower than the portfolio using the passive strategy. Also, obviously, the return derived from arbitrage using ECM is way higher than the passive method (1.06% > 0.075%). With both higher return and lower risk exposed to investors, we may conclude that the ECM portfolio contruction method is very effective.

Conclusion

This paper built an ECM model and find the optimal weights for corn and corn starch futures contracts are 1:1.37425. It means that at this ratio of capital allocation, an investor can obtain the most benefits from it. By building a GARCH (1,1) model, trading strategy is specified as “buy corn starch and sell corn” if the price spread deviate from one unit of its standard deviation, “sell corn starch and buy corn” if it goes below one unit of its standard deviation. Following this trading strategy, excluding trading costs, we get a return of 1.06% from arbitrage, which is higher than that of passive approach, 0.0796746%. This shows our model is effective in making higher return.

Different risk management models have also been exercised to examine its risk. From both EVT and Monte Carlo Simulation approach(Yang et al, 2010)[7], we conclude that the model and trading strategies built and given by ECM and GARCH have both a higher return and a lower risk than the passive investment method.

References

[1] Nikos, S., Nick, K., Georgios, D., 2006. An Intelligent Statistical Arbitrage Trading System. Song, X., 2018. Research on Corn Futures and Corn Starch Futures Arbitrage Based on GARCH Model. Anhui University. Available at:<http://g.wanfangdata.com.cn.ez.xjtlu.edu.cn/index.html> [Accessed on 11th November 2019]

[2] Teng, H., 2017. A spherical Monte Carlo approach for calculating value-at-risk and expected shortfall in financial risk management. 2017 Winter Simulation Conference (WSC), Simulation Conference (WSC), 2017 Winter, p.469. doi: 10.1109/WSC.2017.8247808

[3] Chen,Y., Yu, W., 2019. Setting the margins of Hang Seng Index Futures on different positions using an APARCH-GPD Model based on extreme value theory. Physica A: Statiatical Mechanics and its Applications. doi: 10.1016/j.physa.2019.123207.

[4] Di, H., Zhao, X., Zhang, Z., 2018. Research on Commodity Futures Investment Strategy Based on LSTM-Adaboost Model.

[5] Du, Y., 2018. Application of stress test in commodity futures portfolio – GARCH model based on extreme value theory. East China University of Political Science and Law. Available at:<http://g.wanfangdata.com.cn.ez.xjtlu.edu.cn/index.html> [Accessed on 11th November 2019]

[6] Voropaev, M., 2009. Variance-covariance based risk allocation in credit portfolios: analytical approximation. Available at: http://search.ebscohost.com/login.aspx?direct=true&db=edsarx&AN=edsarx&AN=edsarx.09 05.0781&site=eds-live&scope=site.

[7] Yang, J., Zhang, S., Li, Y., 2010. VaR Model Based on Extreme Value Theory and Risk Management Research.