Study on Statistical Arbitrage in Futures Market

Midterm Presentation

Participant:
Shihong Li

Sponsor:
Greenwoods Asset Management Ltd.

Prepared Date: October 13, 2012

CONTENT

1. Terminology

Arbitrage
The possibility of a risk-free profit at zero cost.

Statistical Arbitrage

1. An investment process based on mathematical models
2. Aiming at making profits
3. Building up long and short positions for assets
4. Taking advantage of asset prices‘ deviation from theoretical values

2. Background Information

Prerequisite:

The securities market in which short selling exists.

Financial Market in China:

1. Before:
   - 1.1 Absent of short selling mechanism
   - 1.2 Absent of stock index futures
2. Now:
   - 2.1 Launched stock index futures on April 16 2010
   - 2.2 Inprovement in startup of securities margin trading

3. Sponsor‘s Background

Greenwoods Asset Management Ltd.

An investment management company specializing in managing investments into mainland China companies. Greenwoods currently manage funds investing in Greater China equities for global investors and A-share trusts for qualified Chinese domestic investors.

4. Deliverables

From team to sponsor:

1. We are going to present an algorithm along with a model in the end of this project.
2. The spread of featured contracts of stock index futures can be predicted.
3. Statistical arbitrage opportunities can be detected by our models.
4. Criteria for entering transactions with arbitrage opportunities can be determined.
5. R packages with a complete set of documents will be created.
6. Technical report and presentation summarizing the work.

4. Deliverables

From sponsor to team:

1. A list of portfolio of interest is needed
2. Sponsor's historic transaction data is needed for modeling, testing, and prediction
3. Computing resources
4. We also expect weekly conference calls for inquiries and consulting

Problem Statement

Data Analysis
To discover arbitrage opportunities, its crucial to extract information from data of historic transactions and featured stock index prices.
We are trying to work out the hidden connection between past data and future trends and make predictions based on this.

Financial Analysis
Based on intensive financial analysis, we seek for benchmarks of arbitrage opportunities for our sponsor.

1. Data Collection

1. Our targeting data should be historic closing prices of contracts of Chinese stock index futures.
2. We choose **CSI300** to be used in our study. CSI 300 is a weighted stock market index designed to replicate the performance of 300 stocks traded in China stock exchanges.

2. Mathematical Models:

Test for Stationarity:

1. Process data before testing
2. Use unit root test to check the property of stationarity

Remark:
a. A stationary process is a stochastic process whose joint probability distribution does not change when shifted in time or space.
b. Providing stationarity of our data, we can apply time series models to analyze data.
b. In statistics, a unit root test tests whether a time series variable is non-stationary using an autoregressive model.

2. Mathematical Models:

Time Series Analysis:

1. Simulate price spreads with AR (1) process model

$$X_t = c + \phi X_t - 1 + \epsilon_t$$

2. Apply GARCH (1,1) model to simulate the residual series of above

$$\epsilon_t = h_t z_t, \quad h_t^2 = \omega + \alpha y_{t-1}^2 + \beta \sigma_{t-1}^2$$

   Here, z_t and ϵ_t are white noise sequences, and other coefficients satisfy basic assumptions of AR(1) and GARCH(1,1) models.

3. Predict future price spreads by plugging real present data into the model.

3. Discover Arbitrage Opportunity

Basic ideas:

1. Conduct financial analysis to calculate cost and profit
2. Predict cash flows of investment
3. Set up criteria for entering and leaving transactions

4. Test and Modification

Basic ideas:

1. Real market data will be plugged into the model in order to test the model.
2. Make modification and improvement to our model.

Progress Report

Accomplishments:

1. Massive data has been collected.
2. Stationarity of data has been verified.
3. AR(1) and GARCH(1,1) models have been applied to our data.

R code

Unit Root Test:

```
library(fUnitRoots)
unitrootTest(cpdiff)
lm.reg=lm(cp1110,cp1109)
summary(lm.reg)
lm.reg.resid=lm.reg$resid
unitrootTest(lm.reg.resid)
```

R code

Time Series Models:

```
library(rugarch)
e=0
pred.series=array()
pred.vol=array()
pred.series=r[1:300]
vol=array()
for (m in 301:len){
e=e+1
garch11=ugarchspec(variance.model = list(model = "fGARCH",
garchOrder = c(1, 1),
submodel="GARCH"), mean.model = list(armaOrder = c(1, 0),
include.mean = FALSE),
distribution.model = "norm")
myfit=ugarchfit(garch11, r[e:(m-1)], out.sample = 0)
forc=ugarchforecast(myfit, n.ahead=1)
pred.series[m]=as.data.frame(forc,which="sigma")[,2]
pred.vol[m]=as.data.frame(forc, which="sigma")[,1]}
```

Conclusion

Outcomes so far:

1. The data of CSI300 collected satisfies stationarity.
2. AR(1) and GARCH(1,1) models are applicable to our data.

Work remaining:

1. Price spreads prediction
2. Financial analysis with respect to sponsor‘s portfolio
3. Model test

Conclusion

Recommendation for future research:

1. It's crucial to select accessible and applicable data carefully.
2. Before setting up mathematical models, one should study theoretical models in depth.

Thank you!