Project Work Statement

Sponsor

Greenwoods Asset Management Ltd.

Study on Statistical Arbitrage in Futures Market

Participants

Shihong Li, sli50@jhu.edu

Supervisor

Nam Lee, nhlee@jhu.edu

Date: October 12, 2012

Any apparent association of this work to Greenwoods Asset Management Ltd. is fictional one, and the sole purpose of this work is a class exercise.

1 Background

First used in 1990s, statistical arbitrage is an investment process based on mathematical model, aiming at making profits by building up long and short positions for assets whose prices deviate from their theoretical values. The technique of statistical arbitrage is to identify statistical mispricing between assets of portfolio, to model the dynamics of this mispricing, to generate the statistical arbitrage strategy and to put it into practice.

On April 16 2010, stock index futures were launched in China. The strategy of statistical arbitrage depends on the securities market in which short selling exists. Being absent of short selling mechanism and stock index futures in Chinese securities market for a long time, statistical arbitrage could not be realized all the time before. However, the startup of securities margin trading and the transaction of **CSI300** futures after the year of 2010 provide a platform for statistical arbitrage. The CSI 300 is a capitalization-weighted stock market index designed to replicate the performance of 300 stocks traded in the Shanghai and Shenzhen stock exchanges. Arbitrage opportunities exist undoubtedly under such inefficient and imperfect market. From this point, research on application of statistical arbitrage to CSI300 futures is very important for Greenwoods Asset Management Ltd..

Greenwoods Asset Management is 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.

2 Problem Statement

To discover arbitrage opportunities, its crucial to extract information from data of historic transactions and featured stock index prices. Currently, Chinese stock index futures appear to be unpredictable and random. Because of the large number of stocks involved, the high portfolio turnover and the fairly small size of the effects we are going to capture, the project is in great need for data mining and statistical models. We are trying to work out the hidden connection between past data and future trends and make predictions based on this. Also, its important and challenging to define criteria for arbitrage opportunity for our sopnsor.

The sponsor currently has a limited capability to make investment decisions given past market information. And our task is to provide them with a reasonable algorithm for detecting potential arbitrage opportunities.

3 Approach

- First, we specify categories of data we need to build our model with. As far as we are concerned, our targeting data should be historic closing prices of contracts of Chinese stock index futures.
- Second, time series models will be carefully chosen to study data relationship and predict future trends. Certain examination must be included to check applicability of time series models.
- Third, criteria for enter a transaction will be set based on mathematical concerns and real market concerns.
- Fourth, real data will be plugged into the model and be tested for validity.
- At last, we make modification and improvement to our model.

4 Milestones

We have the following major deadlines:

- Design project and work statement, Sep 28, 2012,
- Select and collect data, Oct 5, 2012,
- Familiarize ourselves with related software and algorithm to help build models, Oct 10, 2012,
- Build mathematical model and prepare for midterm presentation, Oct 12, 2012,
- Exam data validity, carry empirical test, and modify model if needed, Oct 26, 2012,
- Test model with up-to-date data and prepare for final presentation, Nov 3, 2012,
- Write final report and present results to the sponsor, Nov 30, 2012.

5 Deliverable

5.1 From Team to Sponsor

The following outputs are expected from this project:

- We are going to present an algorithm along with a model in the end of this project,
- The future spread of featured contracts of stock index futures can be predicted,
- Statistical arbitrage opportunities can be detected by our models,
- Criteria for entering transactions and seize arbitrage opportunities can be determined,
- R package with a complete set of documentations along with some test codes that can be used for data analysis, prediction, and test,
- Technical report and presentations summarizing the work.

5.2 From Sponsor to Team

In order for our project to be of successful one, we will need:

- A list of sponsor's main business area in financial market investment, Oct 3, 2012,
- Sponsor's historic transaction data is needed for modeling, testing, and prediction. (However, if it is not available, we can use Bloomberg market data as alternative, Oct 5, 2012,
- Computing resources, Oct 12, 2012,
- Timely We also expect weekly conference calls for inquiries.

References

[1] Andrew Pole, 2007, "Statistical Arbitrage Algorithmic Trading Insights and Techniques," *John Wiley & Sons, Inc..*

[2] Ross, S., 1976, "Options and Efficiency," *Quarterly Journal of Economics, 90.*

[3] Bates, D., 2001, "The Market for Crash Risk," *University of Iowa.*