

# An Dynamic Statistical Arbitrage Trading System

Liu Yang

College of Economics  
Shanghai University  
Shanghai, China

e-mail: liuyang1446@126.com

Lu Guibin

College of Economics  
Shanghai University  
Shanghai, China

e-mail: guibinlu@shu.edu.cn

**Abstract—Objective** – The paper’s aim is to explore an efficient statistical arbitrage system. **Methods** – The paper use moving-window and Regression model to identify the volatility of the relation between two assets. When the relation move beyond normal range which defined by quantile, arbitrage opportunities occur. **Result** –Residuals from moving-window regression model is very close to normal distribution. Generally the arbitrage system is profitable under different parameters. An instance show the system’s total rate of return is 18.9%. **Conclusion** –The profit curve tend to be flat in recent years. Parameters used in the framework should be changed intelligently, because misprice may be corrected in shorter or longer term than history.

**Keywords-component:** *Statistical arbitrage; Mispricing; Quantile; Normal Distribution ; Dynamic-GARCH;*

## I. INTRODUCTION

No risk arbitrage opportunity can be said is the dream of many investors, but as long as the market is free ,then the feasible risk-free arbitrage opportunities hard to exist for a long time; And even there is no-risk arbitrage opportunities, the arbitrage profit will be very modest, not enough to engage the risk-free arbitrage become a continuous work; But we can not deny that the market does sometimes exist some profitable arbitrage opportunities, when the market goes wrong.

Statistical arbitrage is a popular trading strategy employed by hedge funds and proprietary trading desk, built on the statistical notion of cointegration to identify profitable trading opportunities. Statistical arbitrage refers to make profit from pricing deviation in a group of assets. Mispricing is based on the combination of a group of time series which have determined linear relationship with each other, and has a mean reversion and finite variance. For instance, if we have a portfolio  $X_1, \dots, X_n$ , a statistical mispricing can be considered as a linear combination  $\omega = (\omega_1, \omega_2, \dots, \omega_n)$  such that

$$\omega_1 X_1 + \omega_2 X_2 + \dots + \omega_n X_n : \text{mean reverting}(0, \sigma_t^2), \sigma_t^2 < \infty \quad (1)$$

In the above formula we allow variance  $\sigma_t^2$  being dynamic. Vector  $\omega$  actually represents each asset’s weight in the trading portfolios (minus means short

selling corresponding assets). To ensure that the mispricing will eventually „disappear“, we have to make sure the „portfolio“ should be mean reversion. If pricing deviation go beyond our estimation, it will be impossible for us to control the trading risk.

The standard approach to identify statistical mispricing is to run a regression of the values of one asset against the others and test the residuals for mean-reversion. There are some tests can be used for this purpose in the econometric literature. The most common used are the Dickey-Fuller and Phillips-Perron. Note that the residuals of the regression model represent the mispricing at each time  $t$  of  $X_{1t}$  relative to  $\{X_{2t}, \dots, X_{nt}\}$ .

The prediction capability of combination of asset prices tend to be higher the individual. In terms of the ratio of signal and noise, the combination performs better than individual. Apparently, combination improves the predictable components [9]. After we had a synthetic asset, we should create a model that describes the dynamics of mispricing how errors of different magnitude and sign are correct over time. In order to take advantages of predictability, we should use price forecast method into a dynamic trading strategy. An arbitrage trading system identifies the “turning points” of the mispricing time-series and takes proper position on the constituent assets when mispricing become exceptionally high. The arbitrage strategy introduced above is not no-risk. The main risk the strategy faced is that the mispricing has the probability of going beyond expectation we estimated based on history. Generally, the weaker the mean-reversion the higher the probability of observing adverse movements of the synthetic.

There are some research that have suggested approaches that attempt to take advantage of price discrepancies by taking proper transformations of financial time-series; see [6] apply intelligent computation methods such as the feedforward neural network for two Indian stocks. See [7,9] use the idea described above to arbitrage stocks of FTSE 100 Stock Index; [12] for stock index futures and [8] for exchange rate. In [6] Nikos and Nick develop a new arbitrage approach based on popular statistical arbitrage theory, to forecast the volatility of mispricing and established the arbitrage intervals. This paper make some progresses based on research of Nikos and Nick, that we use different length moving window

when we train and trade and also use the latest information in regression model. Without neural network, we built a simpler but more efficient statistical arbitrage framework which compute less when trading and so has high trading speed which is especially important in high frequency trading and also when the price move fast in the market. We calculate the variance of 1000 history samples of mispricing as the variance of mispricing distribution.

The structure of this paper as follows: We describe the data we used in Chapter II. In chapter III we explain the components of the mispricing we made and also the statistical characteristic of it. In chapter IV we give the exactly process to establish MW- regression model and how to use it to forecast the volatility of mispricing in the future. We introduce a trading system which trade based on low frequency data in chapter V. In chapter VI we discuss our conclusion and the direction of future research.

## II. SAMPLE DATA

This paper selects the price data of Ping An Bank and SPD Bank in Shenzhen Stock Exchange and Shanghai Stock Exchange respectively. The reasons why we chose those two stocks are as follows:

- These two stocks have very long history in Chinese stocks market. We can use them to test the long term stability of the arbitrage framework this paper introduced.
- These two stocks have a strong complementarity in the market, so as their prices correlation. In terms of property of mean reverting, the synthetic spreads of these two assets will have a better performance.
- Contrast to Europe and the United State, China is an emerging market which is easily influenced by external fluctuation. It will be not efficient if we use some strategy that we used in mature market to arbitrage in such market. So this paper intend to develop a new arbitrage framework that can be used in emerging market like China.

## III. IDENTIFYING STATISTICAL MISPRICINGS

Ping An and SPD are two of the biggest market value bank. We used forward answer authority price of these two stocks to generate a synthetic asset which actually is the mispricing. Fig .1 shows daily closing price of Ping An and SPD from November 10,1999 to October 28,2014. Table I shows the correlation coefficients between some stocks. As we can see that Ping An and SPD have a relatively high correlation with each other.

TABLE I. CORRELATION COEFFICIENTS OF SOME PAIRS OF STOCKS

	PING AN	SPD	HXB	CMBC	CMB
PING AN	1	0.966	0.951	0.874	0.926
SPD	0.966	1	0.961	0.895	0.946
HXB	0.951	0.961	1	0.835	0.939
CMBC	0.874	0.895	0.835	1	0.812
CMB	0.926	0.946	0.939	0.812	1

a. Daily data from 03/01/2000 to 29/10/2014

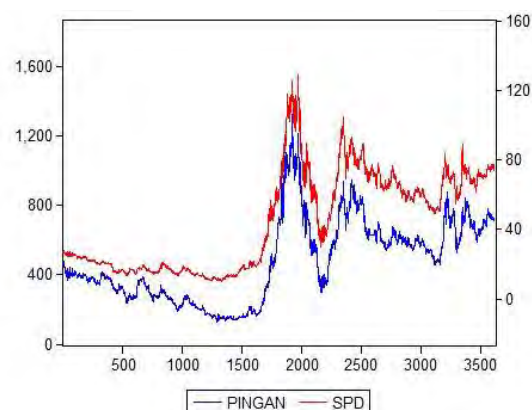


Figure 1. Daily data of PA and SPD from 10/11/1999 to 28/10/2014

As a first attempt to construct a synthetic asset, we ran a regression of SPD against Ping An, hence forth  $x_2$  and  $x_1$  respectively, over all the observations and we then got the residuals (which can be considered as mispricing) of the regression model. The resulting series is depicted in Fig .2 and Fig .3. Observe that the synthetic asset is weakly mean-reverting. But the PP test reject the hypothesis of the residuals is non-stationary( with p-value=0.01), while normal test reject the hypothesis of the residuals is normal distribution(with P value<2.2e-16). It implies that although the pair of stocks have strong correlation ship, the statistic regression model between them is not stable over time.

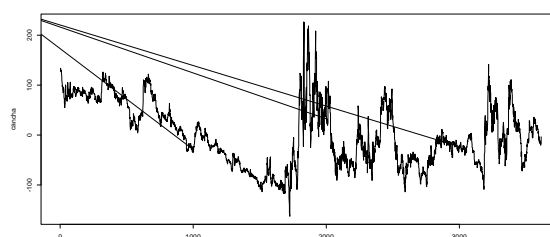


Figure 2. The synthetic asset constructed from a static regression

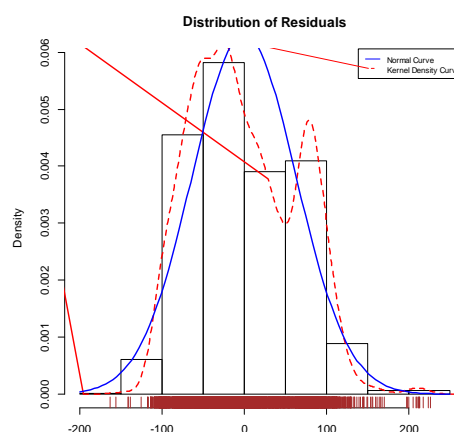


Figure 3. Distribution of the synthetic asset (3618 observations)

In order to control the non-mean reverting of the synthetic asset, we adopt an dynamic transformation scheme in which the coefficients of the linear combination are periodically re-calculated. In particular, we define the mispricing as

$$Z_t = X_{2,t} - \alpha_{t-1} - \beta_{t-1}X_{1,t} \quad (2)$$

Where  $\alpha_t, \beta_t$  are estimated on the basis of a window of length  $W$ . In [6], Nikos and Nick provide an approach which is slightly different from linear regression. It defines  $\beta$  as the mean price ratio between the two stocks over the specified window and subsequently choose  $\alpha$  so as to minimize the total variation of  $Z_t$  within the window, i.e.

$$\beta_t = \text{mean}(X_{2,j}/X_{1,j}, j=t-W+1, \dots, t) \quad (3)$$

$$\alpha_t = \text{mean}(X_{2,j}, j=t-W+1, \dots, t) - \beta_t \text{mean}(X_{1,j}, j=t-W+1, \dots, t) \quad (4)$$

This procedure has been experimentally found to give more reasonable estimates of the synthetic vector that also show more stability over time. In Fig .4 we show the syntetic time series resulting from training window with length 50 and trade window with length 30. In detail , we use the past 50 sample observations to ran regression model to estimate the coefficients of the regression model, then we use the coefficients on the next 30 sample observations to get mispricing. Now we can see that the mispricing is fitting normal curve better than Fig .3.

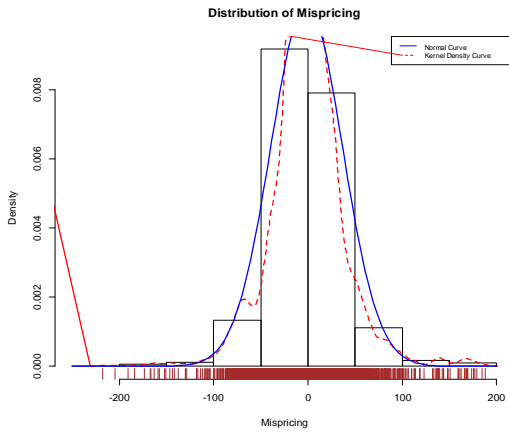


Figure 4. The distribution of mispricing

#### IV. APPLICATION

Our methodology for exploiting statistical arbitrages consists of the following steps:

- Constructing a “synthetic asset” . Synthetics are calculated for various sampling frequencies.
- Forecast the mispricing volatility's range.
- Implementing a trading system to exploit the predictable component of mispricing dynamics. The trading strategy is roughly as follows: buy (long) the synthetic asset if  $Z_t < \hat{Z}_{t+1}^{L,\alpha}$  and sell (short) the synthetic asset if  $Z_t > \hat{Z}_{t+1}^{H,\alpha}$ , where  $\hat{Z}_{t+1}^{L,\alpha}$  and

$\hat{Z}_{t+1}^{H,\alpha}$  denote the  $\frac{\alpha}{2}\%$  quantile and the  $(100-\frac{\alpha}{2})\%$  quantile on the value of the mispricing 1steps ahead. In our approach the confidence interval is a decision variable, which has to be adjusted so that profit is maximized.

Fig .5 reveals the accumulated profit of the arbitrage system. The training window is 50 and the trading window is 30, we use 400 history mispricing to estimate the variance of mispricing. During the past 14 years the system have traded 61 times. The average profit per trade is 11.07, with 0.05% fee ratio and 0.066% (annualized) securities lending fees the cost per trade is 3.16. The net profit per trade is 7.19. With principal of 2553.2, the total rate of return is 18.9%. Under given confidence level, performance of this system rely on the length of training window and trading window selected. Fig .6 shows how profits is changed by different combination of training and trading window under 0.1 confidence level.

In Fig .7 shows the performance of arbitrage trading strategies by varying the confidence level . Observe that as the interval gets narrower( confidence level increased) the net accumulated profit becomes higher at first but then fluctuate and decreases as the costs become higher while the trading time increase. Fig .8 shows a trading instances of a system with bounds set at 10% confidence.

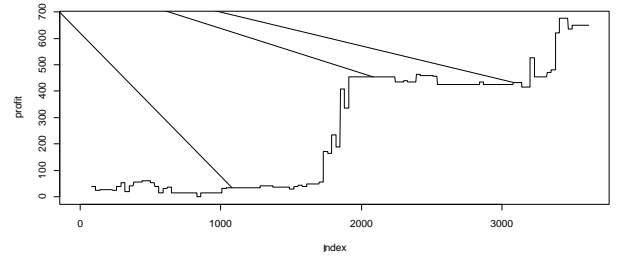


Figure 5. The accumulate profit curve

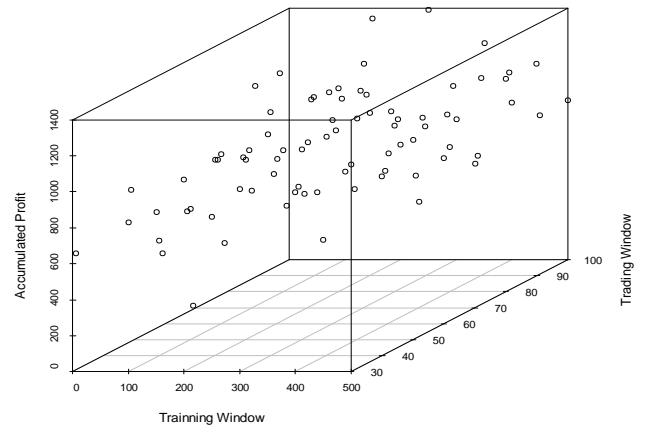


Figure 6. Performace of the system under diferent window length

a. under 0.1 confidence level

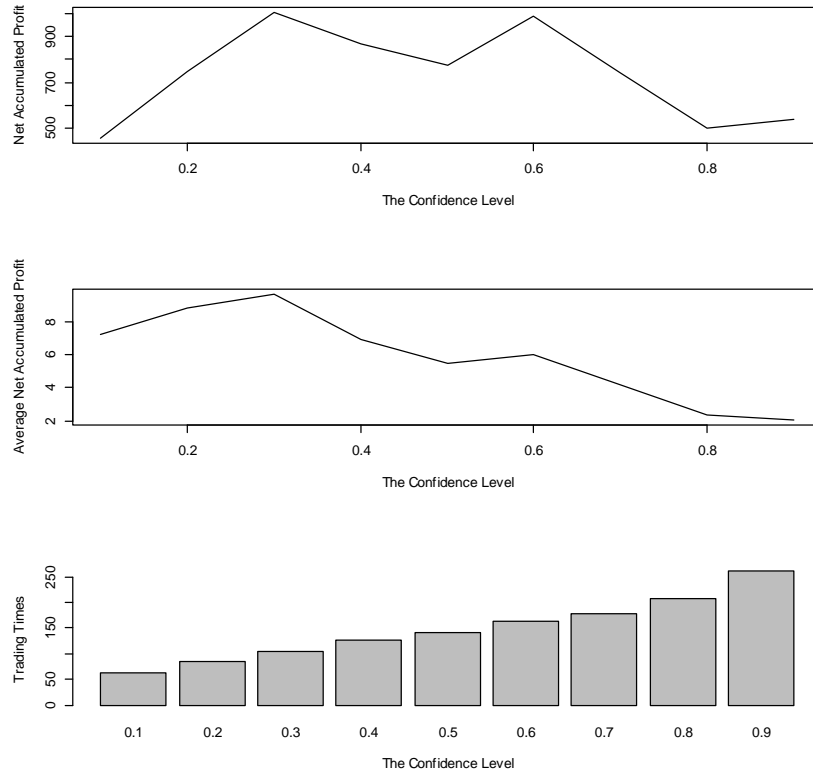


Figure 7. Performace of the system under diferent confidence level

a. training window is 50 while trading window is 30

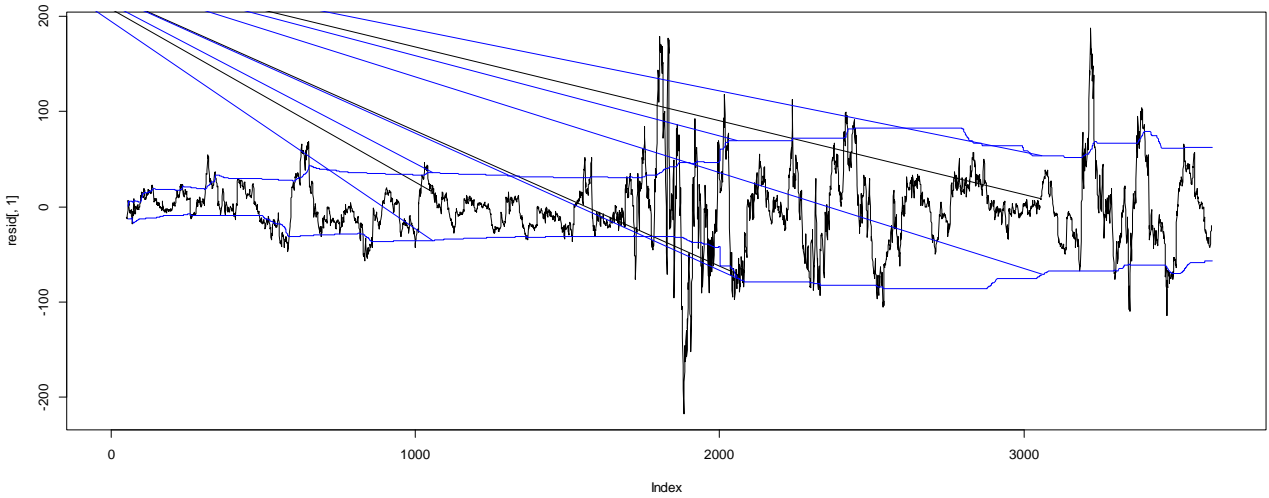


Figure 8. Trading instance

a. Blue lines represent a 10% confidence "envelop" for the value of the mispricing.

## V. DISCUSSION –FURTHER RESEARCH

This paper introduced a new computational intelligent framework for detecting and exploiting statistical arbitrage opportunities in a group of assets. Experiments show that there are great profits especially in

the Chinese stocks market, so that the frame has good application significance. Unlike other intelligent approaches, we do not focus on point forecasts but on the volatility to estimate the fluctuation range in a certain probability of the future.

In this paper, the results show that the system have a good performance in mining arbitrage opportunities and profitability. Actually, there are still many problems in actual application. The profitability of these trades in a real market environment is still questionable given the various trading costs and market “frictions”. At present, we conducting further research on the tuning of the parameters of our trading system, and applying this arbitrage system on high frequency trading.

#### REFERENCES

- [1] Andrew Neil Burgess: A Computational Methodology for Modelling the Dynamics of Statistical Arbitrage[D], University of London for the degree of Doctor of Philosophy, 1999.
- [2] E. Steurer and T.H Hann: Exchange rate forecasting comparison: neural networks, machine learning and linear models [A]. Neural Networks in Financial Engineering (A. Refenes, Y. Abu-Mostafa, J. Moody, and A. Weigend, eds.)(C], World Scientific, 1996, pp. 113-121.
- [3] I. Garrett and N. Taylor: Intraday and interday basis dynamics: Evidence from the FTSE 100 index futures market [J]. Studies in Nonlinear Dynamics and Econometrics 5 (2001), no.2,133-152(20).
- [4] Mihir Dash, Jaykumar H. Dagha: An Application of GARCH Models in Detecting Systematic Bias in Options Pricing and Determining Arbitrage in Options [J]. Journal of CENTRUM Cathedra. Vol 5, Issue 1, 2012. 91-101.
- [5] N. Burgess: Statistical arbitrage models of the FTSE 100[A]. Computational Finance 1999 (Y. Abu-Mostafa, B. LeBaron, A. W Lo, and A. S. Weigend, eds.)(C], The MIT Press, 2000, pp.297-312.
- [6] Nikos S. Thomaidis, Nick Kondakis: An intelligent statistical arbitrage trading system[A] . Advances in Artificial Intelligence[C], Lecture Notes in Computer Science Volume 3955, 2006, pp 596-599
- [7] Norman R. Swanson, Tian Zeng: Choosing Among Competing Econometric Forecasts: Regression-based Forecast Combination Using Model Selection[J]. Journal of Forecasting,425-440(2001).
- [8] N.S. Thomaidis, G. Dounias: Financial statistical modelling with a new nature-inspired technique[A]. in Proc. Of the 1st European Symposium on Nature-Inspired Smart Information Systems (NISIS)(C], Albufeira, Portugal,2005.
- [9] N. Towers, Cointegration: Neural Networks and the Financial Markets: predicting, combining and portfolio optimization [A]. (J. Shadbolt and J. G. Taylor, eds.)(C], Spriger, 2002,pp. 181-201.
- [10] Richard Roll, Stephen A. Ross: An Empirical Investigation of the Arbitrage Pricing Theory[J]. The Journal of Finance, Vol. 35, No. 5(Dec., 1980), 1073-1103.
- [11] Robert I. Kabacoff: R in Action: Data Analysis and Graphics with R[M]. Posts & Telecom Press, 2011
- [12] Thomas A. Hanson, Joshua R. Hall: Statistical Arbitrage Trading Strategies and High Frequency Trading, January 12, 2013.