Review of Pairs Trading Strategies[1]

Kuan-Lun Wang

polyphonicared.am04g@g2.nctu.edu.tw,
National Chiao Tung University

Research Group Meeting

[1] The talk is based on the work of Krauss (2016). [12]

Outline for Review of Pairs Trading Strategies

Background: Mathematical Review

Definition: Random Variable [10]

Given a random experiment with an outcome space S, a function X that assigns one and only one real number $X(s) = x$ to each element s in S is called a *random variable*. The *space* of X is the set of real numbers $\{x : X(s) = x, s \in S\}$, where $s \in S$ means that the element s belong to the set S.

Definition: Random Process [13]

Let $\{\Omega, \mathcal{F}, P\}$ be a probability space and $T = [0, \infty)$. The family $X = (\xi_t)$, $t \in T$, of random variables $\xi_t = \xi_t(\omega)$ is called a (*real*) *random process with continuous time* $t \in T$. In the case where the time parameter t is confined to the set $\mathbb{N} = \{0, 1, \ldots, \}$, the family $X = (\xi_t)$, $t \in \mathbb{N}$, is called a random sequence or a random process with discrete time.

Background PTS Distance Cointegration References
Background: Mathematical Review

Background: Mathematical Review

Definition: Arbitrage [15]

An *arbitrage* is a portfolio value process $X(t)$ satisfying $X(0) = 0$ and also satisfying for some time $T > 0$

$$\Pr[X(T) \geq 0] = 1, \Pr[X(T) > 0] > 0.$$

Definition: Statistical Arbitrage [15]

A *statistical arbitrage* is a portfolio value process $X(t)$ satisfying for some $\epsilon > 0$, for some time $T > 0$

$$\Pr[X(T) - X(0) \geq 0] > 1 - \epsilon.$$

Notation: Arbitrage vs. Statistical Arbitrage

To be more specific, if a portfolio value process is an arbitrage, then it must be statical arbitrage. Conversely, not all statistical arbitrages are arbitrages.

Pairs Trading Strategies

According to Gatev *et al.* (2006), the concept of pairs trading is surprisingly simple and follows a two-step process:

Step 1. find two securities whose prices have moved together historically in a formation period.

Step 2. monitor the spread between them in a subsequent trading period.

- If the prices diverge and the spread widens, short the winner and buy the loser.

Pairs Trading Strategies

The concept of univariate pairs trading can be extended:

Frameworks 1. Quasi-Multivariate: one security is traded against a weighted portfolio of comoving securities.

Frameworks 2. Multivariate: groups of stocks are traded against other groups of stocks.

Terms of reference for such strategies are (quasi-)multivariate pairs trading, generalized pairs trading, or statistical arbitrage.

The most cited paper in this domain is published by Gatev *et al.* (2006), hereafter GGR.

Pairs Trading Strategies

Table: Overview Pairs Trading Approaches

Approach	Representative Studies	Sample	Return p.a.
Distance	Gatev *et al.* (2006)	U.S. CRSP 1962-2002	0.11
	Do and Faff (2010)	U.S. CRSP 1962-2009	0.07
Co-integration	Vidyamurthy (2004)	-	-
	Rad *et al.* (2015)	U.S. CRSP 1962-2014	0.10
Time series	Elliott *et al.* (2015)	-	-
	Cummins and Bucca (2012)	Engrgy futures 2003-2010	≥ 0.18
Stochastic	Jurek and Yang (2007)	Selected stocks 1962-2004	0.28-0.43
control	Liu and Timmermann (2013)	Selected stocks 2006-2012	0.06-0.23
Others: ML, combined	Huck (2009)	U.S. S&P 100 1992-2006	0.13-0.57
forecasts	Huck (2010)	U.S. S&P 100 1993-2006	0.16-0.38
Others:	Krauss and St ubinger (2015)	U.S. S&P 100 1990-2014	0.07-0.08
Copula	Rad *et al.* (2015)	U.S. CRSP 1962-2014	0.05
Others: PCA	Avellaneda and Lee (2010)	U.S. subset 1997-2007	-

The Baseline Approach

Gatev *et al.* (2006) study is performed on all liquid U.S. CRSP stocks from 1962 to 2002.

Step 1. a cumulative total return index is constructed for each stock and normalized to the first day of a 12-month formation period.

Step 2. with n stocks under consideration, the sum of Euclidean squared distance (SSD) for the price time series of $n(n-1)/2$ possible combinations of pairs is calculated.

- The top 20 pairs with minimum historic SSD are considered in a subsequent six-month trading period.
- Trades are opened when the spread diverges by more than two historical standard deviations and closed upon mean-reversion, at the end of the trading period, or upon delisting.

The Baseline Approach

Spread Variance. $p_{i,t}$ and $p_{j,t}$ denote realizations of the normalized price processes $P_i = (P_{i,t})_{t\in T}$ and $P_j = (P_{j,t})_{t\in T}$ of the securities i and j of a pair and $s^2(\cdot)$ sample variance. Empirical spread variance $s^2_{P_i-P_j}$ can be expressed as

$$s^2_{P_i-P_j} = \frac{1}{T}\sum_{t=1}^{T}(p_{i,t}-p_{j,t})^2 - \left(\frac{1}{T}\sum_{t=1}^{T}(p_{i,t}-p_{j,t})\right)^2 .$$

We can solve for the average sum of squared distances $\overline{ssd}_{P_i,P_j}$ in the formation period:

$$\overline{ssd}_{P_i,P_j} = \frac{1}{T}\sum_{t=1}^{T}(p_{i,t}-p_{j,t})^2 = s^2_{P_i-P_j} + \left(\frac{1}{T}\sum_{t=1}^{T}(p_{i,t}-p_{j,t})\right)^2 .$$

Equation shows that constraining for low SSD is the same as minimizing the sum of spread variance and squared spread mean. However, the profit-maximizing rational investor seeks out pairs with high spread variance and strong mean-reversion properties.

The Baseline Approach

GGR interpret the pair's price time series as cointegrated in the sense of Bossaerts (1988). However, the author develops a rigorous cointegration test based on canonical correlation analysis. Conversely, GGR perform no cointegration testing whatsoever (Galenko *et al.*, 2012). As such, the high correlation may well be spurious, since high correlation is not related to a cointegration relationship (Alexander, 2001). These assertions are confirmed in a recent comparison study of Huck and Afawubo (2015), showing that cointegrated pairs are more profitable and their spread volatility twice as high compared to distance pairs.

Expanding on the GGR Sample

Chone and Frazzini (2008)

Observe 1. find substantial customer-supplier links in the U.S. stock market allowing for return predictability.

Observe 2. pairs with a high number of zero-crossings in the formation period are favored.

- This heuristic is used as a proxy for mean-reversion strength, in lieu of cointegration testing.
- The top portfolios incorporating industry restrictions, the number of zero-crossings as well as SSD in the selection algorithm are still slightly profitable, even after trading costs.

From SSD to Pearson Correlation and Quasi-Multivariate Pairs Trading

Chen *et al.* (2012) use the same data set and time frame as GGR, but opt for Pearson correlation on return level for identifying pairs.

Development of a Theoretical Framework

Vidyamurthy (2004) provides the most cited work for cointegration-based pairs trading. The framework relies on three key steps:

Step 1. pairs are preselected based on statistical or fundamental similarity measures.

Step 2. tradability is assessed, following an adapted version of the Engle–Granger cointegration test.

Step 3. optimal entry/exit thresholds are designed with nonparametric methods.

A Large-Scale Empirical Applications

Rad *et al.* (2015) provide the first large-scale empirical application of the cointegration approach on U.S. CRSP data from 1962 until 2014, following GGR and Vidyamurthy (2004). Following GGR:

Step 1. they identify the stocks with minimum SSD in a 12-month formation period.

Step 2. they test these stocks for cointegration with the Engle-Granger approach and only retain the top 20 stocks of the SSD ranking that are also cointegrated.

As such, the similarity in return behavior between distance and cointegration approach comes as no surprise.

Active Statistical Arbitrage Strategies

Whereas correlation reflects short-term linear dependence in returns, cointegration models long-term dependencies in prices (Alexander, 2001). As such, compared to the distance methods, the cointegration approach has a higher potential of identifying true long-term equilibrium relationships between several assets.

Background PTS Distance Cointegration References

Adjacent Developments

Cheng *et al.* (2011) develop a statistical arbitrage strategy with a cointegration model based on logistic mixture autoregressive equilibrium errors.

Kuan-Lun Wang Review of Pairs Trading Strategies

References

[1] Alexander, Carol, and Anca Dimitriu. "Indexing and Statistical Arbitrage." The Journal of Portfolio Management Portfolio Management 31.2 (2005): 50-63. Institutional Investor, Inc.'s Journals. Web. 24 July 2016.

[2] Alexander, Carol. Market Models: A Guide to Financial Data Analysis. 1st ed. UK and New York: Wiley, 2001. Print.

[3] Bossaerts, Peter. "Common Nonstationary Components of Asset Prices." *Journal of Economic Dynamics and Control* 12.2-3 (1988): 347-64. *ScienceDirect*. Web. 24 July 2016.

[4] Chen, Huafeng Jason, Shaojun Jason Chen, and Feng Li. *Empirical Investigation of an Equity Pairs Trading Strategy*. Working paper. N.p.: n.p., 2012. Print.

References

[5] Cheng, Xixin, Philip L.H. Yu, and W.k. Li. "Basket Trading under Co-integration with the Logistic Mixture Autoregressive Model." *Quantitative Finance* 11.9 (2011): 1407-419. Taylor & Francis Online. Web. 24 July 2016.

[6] Cohen, Lauren, and Andrea Frazzini. "Economic Links and Predictable Returns." The Journal of Finance 63.4 (2008): 1977-2011. Wiley Online Library. Web. 24 July 2016.

[7] Do, Binh, and Robert Faff. "Does Simple Pairs Trading Still Work?" *Financial Analysts Journal* 66.4 (2010): 83-95. CFA Institute. Web. 24 July 2016.

[8] Galenko, Alexander, Elmira Popova, and Ivilina Popova. "Trading in the Presence of Cointegration." The Journal of Alternative Investments 15 (2012): 85-97. Institutional Investor, Inc.'s Journals. Web. 24 July 2016.

References

[9] Gatev, Evan, William N. Goetzmann, and K. Geert Rouwenhorst. "Pairs Trading: Performance of a Relative-Value Arbitrage Rule." *The Review of Financial Studies* 19.3 (2006): 797-827. *Social Science Research Network*. Web. 24 July 2016.

[10] Hogg, Robert V., Elliot A. Tanis, and Dale L. Zimmerman. *Probability and Statistical Inference*. 9th ed. Taipei: Pearson, 2015. Print.

[11] Huck, Nicolas, and Komivi Afawubo. "Pairs Trading and Selection Methods: Is Cointegration Superior?" *Applied Economics* 47.6 (2015): 599-613. *Taylor & Francis Online*. Web. 24 July 2016.

References

[12] Krauss, Christopher. "Statistical Arbitrage Pairs Trading Strategies: Review and Outlook." *Journal of Economic Surveys* (2016): n. pag. *Wiley Online Library*. Web. 24 July 2016.

[13] Liptser, Robert Shevilevich., and Albert Nikolaevitch Chiriaev. *Statistics of Random Processes*. 2nd ed. Vol. 5. Berlin Heidelberg: Springer, 2001. Print. Applications of Mathematics.

[14] Rad, Hossein, Rand Kwong Yew Low, and Robert W. Faff. *The Profitability of Pairs Trading Strategies: Distance, Cointegration, and Copula Methods*. Working paper. N.p.: n.p., n.d. *Social Science Research Network*. Web. 24 July 2016.

References

[15] Shreve, Steven E. *Stochastic Calculus for Finance II*. 1st ed. New York: Springer, 2004. Print. Springer Finance Textbooks.

[16] Vidyamurthy, Ganapathy. *Pairs Trading: Quantitative Methods and Analysis*. Hoboken, New Jersey: John Wiley & Sons, 2004. Print.