Journal of Business and Social Science Review
Issue: Vol. 2; No.3; March 2021 pp.1-13
ISSN 2690-0866(Print) 2690-0874 (Online)
Website: www.jbssrnet.com
E-mail: editor@jbssrnet.com
Doi: 10.48150/jbssr.v2no3.2021.a1

APPLICATION OF AN OPTIMAL PAIRS TRADING STRATEGY UNDER GEOMETRIC BROWNIAN MOTIONS AND COINTEGRATION TEST TO THE S&P100

Dong-Hoon Shin

Department of Global Finance and Banking,
Inha University
Incheon, South Korea
Email: dhshin@inha.ac.kr

DongwookKim

Department of Mathematics
Texas A&M University-Kingsville
Kingsville, the United States of America
Email: Dongwook.Kim@tamuk.edu

ChangkiKim

Korea University Business School
Korea University
Seoul, South Korea
Email: changki@korea.ac.kr

Abstract

This paper is concerned with applying an optimal pairs trading strategy to stocks under Geometric Brownian Motions proposed by Tie et al.(2018) to actual market data. Traditional pairs trading strategy usually assume a difference of the stock prices follows a mean-reversion. In addition, a pair selection process is necessary to confirm that the difference between stocks constituting a pair follows a mean-reversion process, and a cointegration test is generally used for pair selection. Compared to the traditional pair trading strategy, the advantage of this strategy is that it allows the stocks used in the strategy to follow general Geometric Brownian motions. In this paper, the optimal pairs trading strategy was applied to the daily closing price data of 83 stocks composing the S&P100 for 20 years. In particular, the cointegration testwas applied for pairs selection. As the significance level of the cointegration test is closer to 0, the difference between the stocks making up pairs is more likely to follow the mean-reversion model. Conversely, as the significance level is closer to 1, the relationship of the stocks making up pairs is free. As the result, it was found that the cointegration test was not necessary for pair selection to improve the overall performance. We also considered conditions of other variables in order to obtain better performance of the strategy.If the variables are well selected, it was confirmed that by applying this optimal pairs trading strategy, a stable annual average return of 14% or more can be obtained.

Keywords: pairs trading, cointegration,optimaltrading strategy, S&P100;

Subject classification codes: 93E20, 91G80, 49L20

1. Introduction

Pairs trading is a trading strategy to track the price movements of two stocks and compare the relative price strengths. In general, pairs trading consists of a short position in a strong stock and a long position in the weak one. This strategy bets on the reversal of their price strength. Pairstrading is attractive to investors due to their ``market neutral" nature, by which traders can make a profit from any market condition. Pairs trading was developed by the Wall Street quant Nunzio Tartaglia at Morgan Stanley; SeeGatev, Goetzmann and Rouwenhorst(2006) for details.

There has been a lot of in-depth discussion regarding the causes of divergence and subsequent convergence; see Vidyamurthy'sbook(2004).Mashele, Terblanche and Venter(2013) suggested formal trading rules of pairs trading strategies by back-testing on stocks listed on the Johannesburg Stock Exchange.

Mathematical approaches of trading rules and portfolio selection have been developed for many years. Zhang(2001) presented an approach to find a target price and a stop-loss limit, obtained by solving a set of two-point boundary value problems, using the regime switching model. Guo and Zhang(2005) derived optimal selling rules under a regime switching Geometric Brownian Motion and proposed that there are optimal threshold levels in trading of stocks that can be obtained by solving a set of algebraic equations with a smooth-fit technique. Dai, Zhang and Zhu(2010) studied a trend following a trading rule based on a conditional probability of a bull market. They derived the optimal trading obtained by two threshold curves, which can be determined by solving the associated Hamilton–Jacobi–Bellman (HJB) equations. Song and Zhang (2013) developed an advanced mathematical method in pairs trading when the portfolio of the underlying pairs follows a mean-reversion model and proposed that the optimal pairs trading rule can be determined by threshold levels of both buying and selling obtained by solving algebraic equations. However, the assumption of the mean reversion model for stock pairs in this research shows a limitation on its applications in selecting stock pairs, which are required to be from the same industrial sector. To relax this constraint, Tie, Zhang and Zhang(2018) developed a general approach of pairs trading under Geometric Brownian Motions that can be used to trade any pairs of stocks without reference to price correlation condition. They presented a sophisticated general approach for pairs trading using two stocks (Walmart and Target) with numerical results. However, there is a lack of efficiency in the performance of their approach in the real market.In this paper, we investigated the pairs trading performance, presented in Tieet al.(2018), with the stocks that make up S&P 100 to understand how the results are realized in the real market.

To improve the efficiency and performance of pair trading, many researchers have applied a cointegration method. Lin, McCrae and Gulati(2006) used the cointegration method to protect the pair trading strategy from serious losses. They applied the OLS method to create spreads and set various conditions that translate into trading behavior. In their models, they have achieved their trading strategy with a minimum level of return that is protected from the risk of loss. The result was an excess return of about 11% per year for the entire period. Mikkelsen(2018) compared the performances of distance and cointegration approaches with using each high-frequency and daily dataset to check whether it is profitable for Norwegian seafood companies and obtained a similar result for the two approaches. Fallahpour, Hakimian, Taheri and Ramezanifar (2016) applied a cointegration test to various pairs of stocks and a vector error-correction model to make a trading signal.

The main contributions of this study are summarized in three ways. First, we modify the model of Tie et al.(2018) with several limitations to fit to the real market and measure the performance of the optimal selling rule using 20 years of historical data for the stocks that make up the S&P100, and verify the effectiveness of the model. Second, since there was no design for cut loss in their work, we try to apply the cointegration test to stock pairs selection to limit the likelihood of heavy losses. In addition, by adjusting the significance level of the cointegration test, we investigate how the pair selection through the cointegration test affects the performance of this strategy. Finally, we confirm the importance of some parameters by analyzing their sensitivity to performance in the real market.

The progress of this study is as follows. In section 2, the problem formulation and notations will be introduced, and the optimal pairs trading strategy under Geometric Brownian Motion obtained from Tie et al.(2018) will be reviewed. Section 3 introduces the data used in this study and describes technical background knowledge for calculating performance that will be obtained using real market data such as cointegration tests, pairs trading strategies, and measure of performance. Section 4 summarizes the algorithm of the strategy, and measures the portfolio's return, equity curves, and Sharpe ratios. In addition, the sensitivities of the rate of return to changes in several variables are investigated. The results are briefly summarized in section 5.

2. The optimal pairs trading strategy under Geometric Brownian Motion and its application

In this section, we review the process of obtaining thresholds for buying and selling stocks through the rule of Tie et al.(2018) and briefly look at what parts to consider for application in the real market. To avoid confusion of notations, many mathematical notations are shared with the notations in their papers.
Let $\{X_t^i, t \geq 0\}$ denote the price of stock S^i for i$\in\{1, 2\}$. Then

$$\mathrm{d}\begin{pmatrix} X_t^1 \\ X_t^2 \end{pmatrix} = \begin{pmatrix} X_t^1 & 0 \\ 0 & X_t^2 \end{pmatrix}\left[\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} dt + \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{pmatrix} d\begin{pmatrix} W_t^1 \\ W_t^2 \end{pmatrix}\right] \quad (1)$$

where μ_i,i∈{1, 2},are the return rates, σ_{ij} , i, j ∈{1, 2}, are the volatility constants, and (W_t^1, W_t^2) is a 2-dimensional standard Brownian motion. Let Z_t mean long S^1 and short S^2at time t. Let l ∈{1, 2} denote the initial net position, and let $\tau_0 < \tau_1 < \tau_2 < \cdots$ denote sequence of stopping time. If initially the net position is long (l = 1), then one should sell $Z_t$before acquiring any future share, i.e., trading sequence is denoted by $\Lambda_1 = (\tau_0, \tau_1, \tau_2, \ldots)$.If the net position is flat (l = 0), then one should start to buy a share of Z_t, i.e., trading sequence is denoted by$\Lambda_0 = (\tau_1, \tau_2, \tau_3 \ldots)$.

Let K denote the fixed percentage of transaction costs associated with buying and selling of stocks S^i,i∈{1, 2}. Let $\beta_b = 1 + K, \beta_s = 1 - K$. Given the initial state (x_1, x_2), the initial net position l ∈{1, 2},and the decision sequences Λ_0 and Λ_1, the corresponding reward functions

$$\begin{aligned} J_0(\mathrm{x}_1, \mathrm{x}_2, \Lambda_0) = E\{&[e^{-\rho\tau_2}(\beta_s X_{\tau_2}^1 - \beta_b X_{\tau_2}^2) I_{\{\tau_2<\infty\}} - e^{-\rho\tau_1}(\beta_b X_{\tau_1}^1 - \beta_s X_{\tau_1}^2) I_{\{\tau_1<\infty\}}] \\ &+ [e^{-\rho\tau_4}(\beta_s X_{\tau_4}^1 - \beta_b X_{\tau_4}^2) I_{\{\tau_4<\infty\}} - e^{-\rho\tau_3}(\beta_b X_{\tau_3}^1 - \beta_s X_{\tau_3}^2) I_{\{\tau_3<\infty\}}] + \cdots\} \\ J_1(\mathrm{x}_1, \mathrm{x}_2, \Lambda_1) = E\{&e^{-\rho\tau_0}(\beta_s X_{\tau_0}^1 - \beta_b X_{\tau_0}^2) I_{\{\tau_0<\infty\}} \\ &+ [e^{-\rho\tau_2}(\beta_s X_{\tau_2}^1 - \beta_b X_{\tau_2}^2) I_{\{\tau_2<\infty\}} - e^{-\rho\tau_1}(\beta_b X_{\tau_1}^1 - \beta_s X_{\tau_1}^2) I_{\{\tau_1<\infty\}}] \\ &+ [e^{-\rho\tau_4}(\beta_s X_{\tau_4}^1 - \beta_b X_{\tau_4}^2) I_{\{\tau_4<\infty\}} - e^{-\rho\tau_3}(\beta_b X_{\tau_3}^1 - \beta_s X_{\tau_3}^2) I_{\{\tau_3<\infty\}}] + \cdots\}, \quad (2) \end{aligned}$$

where $\rho > 0$ is a given discount factor and $I_{\{A\}}$is the indicator function of an event A. For initial net positions l ∈{1, 2},let$V_l(x_1, x_2, \Lambda_l)$ denote the value functions with $(X_0^1, X_0^2) = (x_1, x_2)$. Namely, $V_l(x_1, x_2, \Lambda_l) = sup_{\Lambda_l} J_l(x_1, x_2, \Lambda_l)$.

Assume that $\rho > \max(\mu_1, \mu_2)$. To find the associated HJB equation, let

$$\mathcal{A} = \frac{1}{2}\left\{a_{11}x_1^2\frac{\partial^2}{\partial x_1^2} + 2a_{12}x_1x_2\frac{\partial^2}{\partial x_1 \partial x_2} + a_{22}x_2^2\frac{\partial^2}{\partial x_2^2}\right\} + \mu_1 x_1\frac{\partial}{\partial x_1} + \mu_2 x_2\frac{\partial}{\partial x_2}. \quad (3)$$

where $a_{11} = \sigma_{11}^2 + \sigma_{12}^2, a_{12} = \sigma_{11}\sigma_{21} + \sigma_{12}\sigma_{22}, a_{22} = \sigma_{21}^2 + \sigma_{22}^2$.Then the HJB equations have the form: for $x_1, x_2 > 0$,

$$\begin{aligned} &\min\{\rho v_0(x_1, x_2) - \mathcal{A}v_0(x_1, x_2), v_0(x_1, x_2) - v_1(x_1, x_2) + \beta_b x_1 - \beta_s x_2\} = 0, \\ &\min\{\rho v_1(x_1, x_2) - \mathcal{A}v_1(x_1, x_2), v_1(x_1, x_2) - v_0(x_1, x_2) - \beta_s x_1 + \beta_b x_2\} = 0. \quad (4) \end{aligned}$$

To convert them into single variable equations, let $y = x_2/x_1$ and $v_l(x_1, x_2) = x_1 w_l(x_1, x_2)$, for some function$w_l(y)$,l ∈{1, 2}. Then

$$\mathcal{A} = x_1\left\{\frac{1}{2}[a_{11} - 2a_{12} + a_{22}]y^2 w_l''(y) + (\mu_2 - \mu_1)y w_l'(y) + \mu_1 w_l(y)\right\} \quad (5)$$

In addition, the HJB equations in terms of $y$and $w_l(y)$can be given as follows:

$$\begin{aligned} &\min\{\rho w_0(y) - \mathcal{L}w_0(y),\ w_0(y) - w_1(y) + \beta_b - \beta_s y\} = 0, \\ &\min\{\rho w_1(y) - \mathcal{L}w_1(y),\ w_1(y) - w_0(y) - \beta_s + \beta_b y\} = 0, \quad (6) \end{aligned}$$

where $\mathcal{L}[w_l(y)] = \lambda y^2 w_l''(y) + (\mu_2 - \mu_1)y w_l'(y) + \mu_1 w_l(y)$and $\lambda = \frac{a_{11} - 2a_{12} + a_{22}}{2}, \lambda \neq 0$. Let

$$\begin{aligned} \delta_1 &= \frac{1}{2}\left(1 + \frac{\mu_1 - \mu_2}{\lambda} + \sqrt{(1 + \frac{\mu_1 - \mu_2}{\lambda})^2 + \frac{4\rho + 4\mu_1}{\lambda}}\right) > 1, \\ \delta_2 &= \frac{1}{2}\left(1 + \frac{\mu_1 - \mu_2}{\lambda} - \sqrt{(1 + \frac{\mu_1 - \mu_2}{\lambda})^2 + \frac{4\rho - 4\mu_1}{\lambda}}\right) < 0. \quad (7) \end{aligned}$$

Then there exists $r_0 > (\frac{\beta_b}{\beta_s})^2$such that $f(r_0) = 0$where

$$\begin{aligned} f(r) &= \delta_1(1 - \delta_2)(\beta_b r^{-\delta_2} - \beta_s)(\beta_b - \beta_s r^{1-\delta_1}) \\ &-\delta_2(1 - \delta_1)(\beta_b r^{-\delta_1} - \beta_s)(\beta_b - \beta_s r^{1-\delta_2}). \quad (8) \end{aligned}$$

With this r_0,$k_1$and $k_2$are given by

$$k_1 = \frac{\delta_2\left(\beta_b r_0^{-\delta_1} - \beta_s\right)}{(1-\delta_2)\left(\beta_b - \beta_s r_0^{1-\delta_1}\right)} = \frac{\delta_1\left(\beta_b r_0^{-\delta_2} - \beta_s\right)}{(1-\delta_1)\left(\beta_b - \beta_s r_0^{1-\delta_2}\right)},$$

$$k_2 = \frac{\delta_2\left(\beta_b r_0^{1-\delta_1} - \beta_s r_0\right)}{(1-\delta_2)\left(\beta_b - \beta_s r_0^{1-\delta_1}\right)} = \frac{\delta_1\left(\beta_b r_0^{1-\delta_2} - \beta_s r_0\right)}{(1-\delta_1)\left(\beta_b - \beta_s r_0^{1-\delta_2}\right)}. \qquad (9)$$

The below two theorems are proved and guarantee the optimality of value functions $V_l(x_1, x_2, \Lambda_l)$.

Theorem 1. Let δ_i be given by (7) and $k_i$be given by (9) for i ∈{1, 2}. Then, the following functions$w_0$and $w_1$satisfy the HJB equation:

$$w_0\,(y) = \begin{cases} \left(\dfrac{\beta_b(1-\delta_2)k_1^{1-\delta_1} + \beta_s\delta_2 k_1^{-\delta_1}}{\delta_1 - \delta_2}\right) y^{\delta_1} & if\ 0 < y < k_2 \\ \left(\dfrac{\beta_b(1-\delta_1)k_1^{1-\delta_2} + \beta_s\delta_1 k_1^{-\delta_2}}{\delta_1 - \delta_2}\right) y^{\delta_2} + \beta_s y - \beta_b & if y \ge k_2 \end{cases}$$

$$w_1\,(y) = \begin{cases} \left(\dfrac{\beta_b(1-\delta_2)k_1^{1-\delta_1} + \beta_s\delta_2 k_1^{-\delta_1}}{\delta_1 - \delta_2}\right) y^{\delta_1} + \beta_s - \beta_b y & if\ 0 < y < k_1 \\ \left(\dfrac{\beta_b(1-\delta_1)k_1^{1-\delta_2} + \beta_s\delta_1 k_1^{-\delta_2}}{\delta_1 - \delta_2}\right) y^{\delta_2} & if y \ge k_1 \end{cases}$$

Define the first quadrant $P = \{(x_1, x_2): x_1 > 0\ and x_2 > 0\}$ and three regions $\Gamma_1 = \{(x_1, x_2) \in P: x_2 \le k_1 x_1\}$,$\Gamma_2 = \{(x_1, x_2) \in P: k_1 x_1 \le x_2 \le k_2 x_1\}$and $\Gamma_3 = \{(x_1, x_2) \in P: x_2 \ge k_2 x_1\}$.

Theorem 2. We have $v_l(x_1, x_2) = x_1 w_l(x_2/x_1) = V_l(x_1, x_2), l = 0, 1$.Moreover, if initially l = 0, let $\Lambda_0^* = (\tau_1^*, \tau_2^*, \tau_3^* \ldots)$ such that $\tau_1^* = inf\{t \ge 0: (X_t^1, X_t^2) \in \Gamma_3\}, \tau_2^* = inf\{t \ge \tau_1^*: (X_t^1, X_t^2) \in \Gamma_1\}, \tau_3^* = inf\{t \ge \tau_2^*: (X_t^1, X_t^2) \in \Gamma_3\}$, and so on. Similarly, if initially l = 1, let $\Lambda_1^* = (\tau_0^*, \tau_1^*, \tau_2^* \ldots)$ such that $\tau_0^* = inf\{t \ge 0: (X_t^1, X_t^2) \in \Gamma_1\}, \tau_1^* = inf\{t \ge \tau_0^*: (X_t^1, X_t^2) \in \Gamma_3\}, \tau_2^* = inf\{t \ge \tau_1^*: (X_t^1, X_t^2) \in \Gamma_1\}$, and so on. Then $\Lambda_0^*$and Λ_1^* are optimal.

Remark 1. In their paper, they mentioned if (k_1, k_2) level for $S^1 - S^2$(long S^1 and short S^2) pair trading was obtained, then $(\widetilde{k_1}, \widetilde{k_2})$ level for $S^2 - S^1$ (long S^2 and short S^1) pair trading was simply given by $\widetilde{k_1} = 1/k_2$,$\widetilde{k_2} = 1/k_1$.

Remark 2. In this paper, the value of the starting position of pair trading consisting of S^i and S^j will be set to zero. That is, $Z_0 = 0$by setting $Z_t = X_t^i/X_0^i - X_t^j/X_0^j$. If the time series $X_t^j/X_t^i$is greater than k_2 at first, we should buy $Z_t$and the performance should follow $V_0\left(x_i, x_j\right)$. If the series $X_t^j/X_t^i$is smaller than k_1 at first, we should sell Z_t and the performance should follow $V_1\left(x_i, x_j\right)$.

3. Data and Technical Background

In this study, 83 stocks from the S&P100 Index in 2020 were selected based on the existence of their data from January 3, 2000, to January 15, 2020(= 5042 observations). To carry out an empirical study, the data must cover the same period, so we discarded 18 stocks from the 101 components[i] of the S&P100 Index.

Table 1 represents the data of stock names and the corresponding abbreviations. We collected adjusted daily closing stock prices using Yahoo financedata[ii].

Table 1. The 83 stocks selected on the S&P Index

No.	Ticker	Company	No.	Ticker	Company
1	AAPL	Apple Inc.	43	INTC	Intel Corp.
2	ABT	Abbott Laboratories	44	JNJ	Johnson & Johnson
3	ADBE	Adobe Inc.	45	JPM	JPMorgan Chase & Co.

4	AIG	American International Group	46	KO	The Coca-Cola Company
5	ALL	Allstate	47	LLY	Eli Lilly and Company
6	AMGN	Amgen Inc.	48	LMT	Lockheed Martin
7	AMT	American Tower	49	LOW	Lowe's
8	AMZN	Amazon.com	50	MCD	McDonald's Corp
9	AXP	American Express	51	MDT	Medtronic plc
10	BA	Boeing Co.	52	MMM	3M Company
11	BAC	Bank of America Corp	53	MO	Altria Group
12	BIIB	Biogen	54	MRK	Merck & Co.
13	BK	The Bank of New York Mellon	55	MS	Morgan Stanley
14	BKNG	Booking Holdings	56	MSFT	Microsoft
15	BLK	BlackRock Inc	57	NEE	NextEra Energy
16	BMY	Bristol-Myers Squibb	58	NKE	Nike, Inc.
17	BRK.B	Berkshire Hathaway	59	NVDA	Nvidia Corporation
18	C	Citigroup Inc	60	ORCL	Oracle Corporation
19	CAT	Caterpillar Inc.	61	OXY	Occidental Petroleum Corp.
20	CL	Colgate-Palmolive	62	PEP	PepsiCo
21	CMCSA	Comcast Corp.	63	PFE	Pfizer Inc
22	COF	Capital One Financial Corp.	64	PG	Procter & Gamble Co
23	COP	ConocoPhillips	65	QCOM	QUALCOMM
24	COST	Costco Wholesale Corp.	66	RTX	Raytheon Technologies
25	CSCO	Cisco Systems	67	SBUX	Starbucks Corp.
26	CVS	CVS Health	68	SLB	Schlumberger
27	CVX	Chevron Corporation	69	SO	Southern Company
28	DD	DuPont de Nemours Inc	70	SPG	Simon Property Group, Inc.
29	DHR	Danaher Corporation	71	T	AT&T Inc
30	DIS	The Walt Disney Company	72	TGT	Target Corporation
31	DUK	Duke Energy	73	TMO	Thermo Fisher Scientific
32	EMR	Emerson Electric Co.	74	TXN	Texas Instruments
33	EXC	Exelon	75	UNH	UnitedHealth Group
34	F	Ford Motor Company	76	UNP	Union Pacific Corporation
35	FDX	FedEx	77	UPS	United Parcel Service
36	GD	General Dynamics	78	USB	U.S. Bancorp
37	GE	General Electric	79	VZ	Verizon Communications
38	GILD	Gilead Sciences	80	WBA	Walgreens Boots Alliance
39	GS	Goldman Sachs	81	WFC	Wells Fargo
40	HD	Home Depot	82	WMT	Walmart
41	HON	Honeywell	83	XOM	Exxon Mobil Corp.

42	IBM	International Business Machines		

3.1 Determination of pairs using the cointegration test

There are several methods for pair selection in pairs trading such as the stochastic approach (Mudchanatongsuk, Primbs and Wong(2008), Tourin and Yan(2013)), the discrete approach (Nath(2004), Rad, Low and Faff(2016)), and the cointegration approach (Fallahpour et al.(2016), Rad et al.(2016). We use the cointegration approach in this article. To introduce the definition of the cointegration, let x_t and $y_t$have unit roots, and$x_t \sim I(1), y_t \sim I(1)$. If there is a relationship between x_t and $y_t$such that

$$e_t = y_t - \beta_1 - \beta_2 x_t \sim I(0)$$

, then $\sim I(1)$, are said to be cointegrated. Cointegration implies that x_t and $y_t$share similar stochastic trends, and theynever diverse too far from each other. Here,we used the Johansen(1988)'s test to determine whether a pair of two stocks among the 83 stocks from the S&P100 components is cointegrated.

For the optimal pairs trading, we need to determine whether the given pair is cointegrated or notand find the thresholds such as $k_1 and k_2$, and we use data from 252 business days for these calculations. We call this period the observation period. Once the thresholds are obtained, the performance of the optimal pairs trading with the thresholds will be calculated for 126 days following the observation period. We call this period the back-testing period. Denote α as the significance level of Johansen's cointegration test. Denote

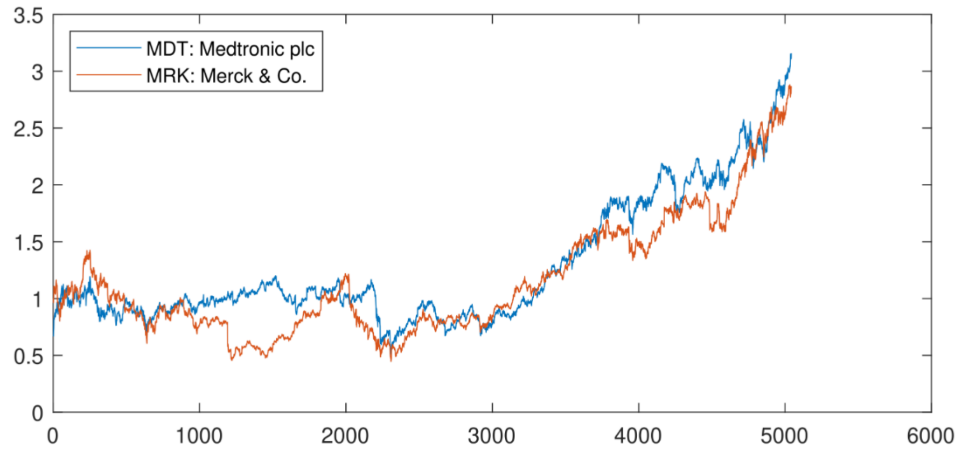
$\mathcal{C}_\alpha^n := \{(i,j)|$ stock i and stock j are cointegrated with significant level αat the period n}

and $|\mathcal{C}_\alpha^n|$as the number of elements of the set $\mathcal{C}_\alpha^n$. Table 2 shows the number of cointegrated pairs on the given observation period of each consecutive period between 2000 and 2019 with $\alpha = 5\%$ significance level on Johansen test among $\binom{83}{2}$= 3403 cases. The period index in the table is the index of the given observation period. If the index is given as n, this period refers to the observation period from January 3, 2000 + 126(n-1) business days to 252 business days thereafter. Figure 1 shows price movements of the pair of cointegrated stocks, Medtronic and Merck & Co., in the sample period from 2000 to 2019.

Table 2. Number of Cointegrated pairs for a given period n. $\alpha = 5\%$.

years	$\lvert\mathcal{C}_\alpha^n\rvert$	years	$\lvert\mathcal{C}_\alpha^n\rvert$
2000	301	2010	88
2001	698	2011	174
2002	495	2012	72
2003	30	2013	53
2004	87	2014	84
2005	142	2015	403
2006	120	2016	319
2007	72	2017	63
2008	203	2018	150
2009	50	2019	95

Figure 1. An example figure of a cointegrated pair: Medtronic vs. Merck & Co.



3.2 Pairs Trading Strategy

The key to pairs trading is to get a signal of when to buy and sell the pair of stocks that have been cointegrated, and section 2 illustrates how to obtain the trading signals (k_1, k_2). When the cointegration is confirmed using the data of the observation period of two stocks i and j, (k_1, k_2) of the two stocks are obtained. Then, using the data of the back-testing period of the two stocks, we find the series of stopping times. If $x_2/x_1 > k_2$ at time $t = \tau^*$, or similarly $\tau^* \in \Gamma_3$, then we take long Z_{τ^*}, and if $x_2/x_1 < k_1$ at time τ^*, or similarly $\tau^* \in \Gamma_1$, then we take short Z_{τ^*}. If the series is Λ_0,$V_0(x_i, x_j)$ is obtained, and if Λ_1,$V_1(x_i, x_j)$is obtained as the performance of the stocks i and j in a period. Define the performance of the pair trading with cointegrated stocks i and j at the period n as

$$\Pi_{ij}^n = \begin{cases} V_0(x_i, x_j) & \text{if the stopping time series} = \Lambda_0 \\ V_1(x_i, x_j) & \text{if the stopping time series} = \Lambda_1 \end{cases}$$

3.3 Measure of Performance

To measure the performance, we need to consider where the principal is invested in Z_t-trading. $Z_t$is a transaction of long one stock $S^1$and selling another stock S^2 at time t. Since the value of the short position is the same as the long position at the beginning of the back-testing, the long position can be financed by the short position because the value of long position stock at the buying time is usually less than or equal to the value of the stock price at the beginning. Namely, the principle is no more used for the long position of Z_t. On the other hand, the short sale requires that the stock S^2 be borrowed at the beginning, and for this purpose, it can be done by depositing other assets or just money of the same value, or at least a fraction of it. This is so-called collateral, and the borrower has to pay an initial margin to his/her margin account. In addition, if the value of the short-selling stock S^2 increases enough, an additional amount must be deposited in the margin account again. This makes it difficult to calculate the investment capital. A more obvious way to calculate the investment capital is to buy the stock S^2 at the beginning of the trading period, and sell S^2 at the end of the trading period where the trading period is from t=0 to t=T. Then, by subtracting $(X_T^2 - X_0^2)/X_0^2$ from the final performance, you can calculate the pure profit/loss Π_{ij}^n for the pairs trading with the investment principal of S^2. We will look at the differences in the performance of these two transactions in section 4.

In addition, the profit/loss formula Π_{ij}^n is sometimes likely to end with the purchase of Z_t at a time t<T. In this case, one should sell Z_T to finish the trading, and the last transaction $Z_t - Z_T$ may be positive or negative. Since it is thought that the expectation of the final calculation will be close to 0 in the mean of the overall calculation, the transaction will be omitted if it ends with the purchase of Z_t at the end of the back-testing period.

If the investor's principal is P, the principal and total profit after all transactions ending in period n can be calculated as $\mathrm{P}\left(1 + \sum_{(i,j)\in\mathcal{C}_\alpha^n} \Pi_{ij}^n / |\mathcal{C}_\alpha^n|\right)$.If the investor rolls over this pairs trading strategy by m times, that is, keeps this contract for m×126 days, the investor's return can be calculated as $\mathrm{P}\prod_{n=1}^{m}\left(1 + \sum_{(i,j)\in\mathcal{C}_\alpha^n} \Pi_{ij}^n / \right.$ $/\mathcal{C}an/$.

4. Data and Technical Background

In this section, we look at the performance of this strategy when we continued the pairs trading strategy with 83 stocks from 2000 to early 2020 (5042 business days). We use the parameters, K = 0.001, α= 5%,andρ =100. The observation period is 1 year (252 days) and the back-testing period is 6 months (126 days) after the observation period. Therefore, there are 38 periods of investing in pairs trading for 6 months. At the end of each period, we will roll over to the next period to see how much return can be obtained over 19 years and compare this with the rate of return of S&P100.The procedure is as follows.

Algorithm:

1. Set K=0.001, α = 0.01, ρ = 10.
2. FOR n=1:38
3. FOR i=2:83
4. FOR j=1:i-1
5. Take daily data set of i, j stocks for the observation period n.
6. IF (i,j) ∈ $\mathcal{C}_\alpha^n$ then
7. Calculate k_1, k_2
8. Take daily data set of i, j stocks for the back testing period n.
9. Calculate Λ_0 or Λ_1
10. Calculate Π_{ij}^n .
11. Calculate the principle and profit for period n = P$\left(1 + \sum_{(i,j)\in\mathcal{C}_\alpha^n} \Pi_{ij}^n / |\mathcal{C}_\alpha^n|\right)$
12. Return Overall Performance=P$\prod_{n=1}^{38}\left(1 + \sum_{(i,j)\in\mathcal{C}_\alpha^n} \Pi_{ij}^n / |\mathcal{C}_\alpha^n|\right)$

Figure 2. An Example of $k_1, k_2, X^2/X^1$, and the equity curve of a pair trading during period 1.

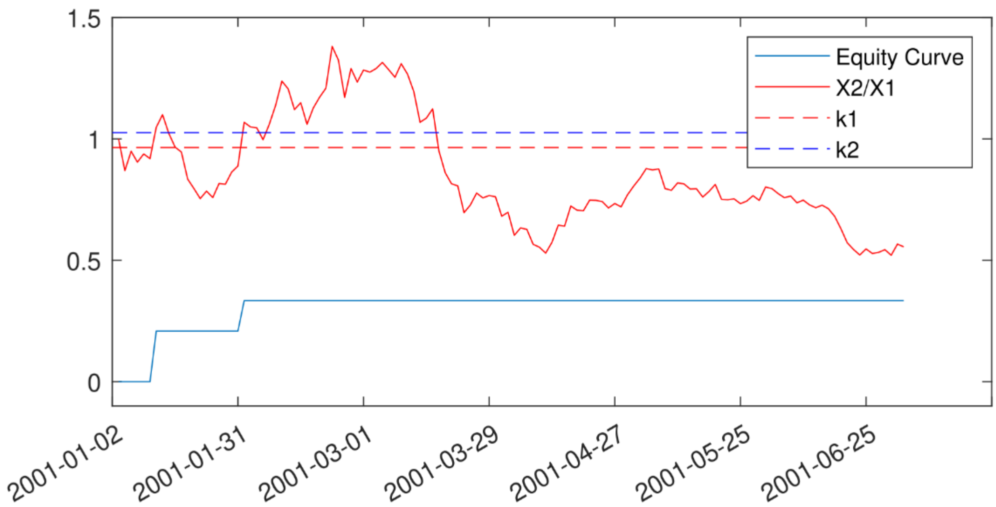


Figure 2 shows the performance of one of the pairs that make up the pairs trading strategy over a period. The observation period for the Figure 2 is the first period, from January 3, 2000 to December 29, 2000, and the back-testing period is from January 2, 2001 to July 2, 2001. S^1 is Adobe Inc., and $S^2$is American Tower. Figure 2 shows the $k_1 = 0.9644, k_2 = 1.0257$ levels calculated during the observation period, the time series graph of X^2/X^1, and the equity curve representing the returns generated by the pair trading.If $100 was invested in this transaction from January 2, 2001 to July 2, 2001, it would be $133.40 as of July 2, 2001, the end of the investment. Of course, this is only one pair with very good performance among the pairs that make up pairs trading, but not all pairs have such good returns. However, by taking the average of the returns of all pairs that are cointegrated, you can obtain the expected return of this strategy.

Figure 3. The equity curve of the pairs trading vs. S\&P100 index in period 1.

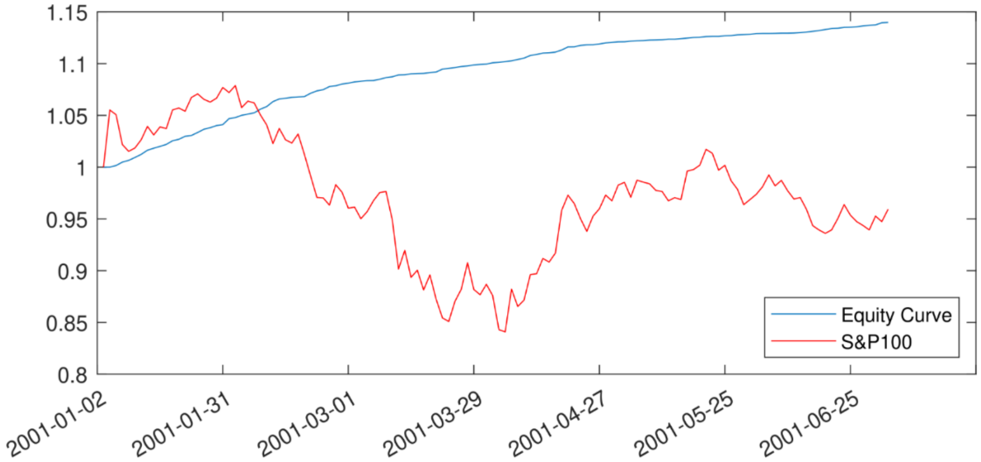


Figure 3 shows the equity curve of the pairs trading consisting of 301 pairs during the back-testing period when the observation period is given from January 3, 2000 to December 29, 2000, and the back-testing period is given from January 2, 2001 to July 2, 2001. It shows that the average return on the pairs trading is 13.98%. For comparison with the S&P100 index, the performance of the S&P100 index at the same period is also displayed.

Figure 4. The equity curves of the pairs trading with or without short selling, and S&P100 index in whole periods.

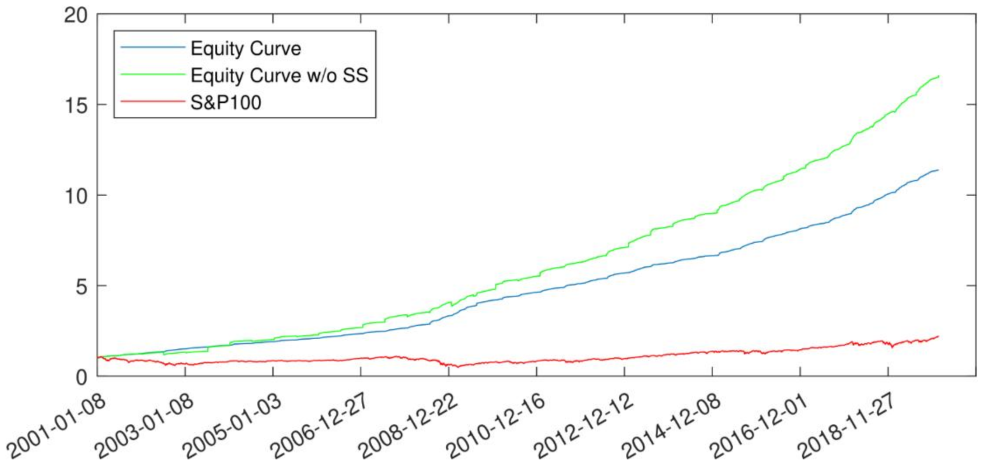


Figure 4 shows the equity curves that can be obtained assuming that the pairs trading strategies with or without short selling are rolled over from 2001 to early 2020, along with the S&P100 index. In the legend of the figure, "Equity curve" represents the equity curve of the pairs trading strategy with short selling, and "Equity curve w/o SS" represents the equity curve of the pairs trading strategy without short selling. In this transaction, we assumed that the pairs trading strategy makes a profit for each period and reinvests the principal and the profit in the pairs trading strategy for the next period until early 2020. In this test, it was found that if an investor invested $100 in principal for about 19 years, the S&P100 index could earn $220.3, whereas the pairs trading strategy could earn more than $1137.95. We also calculated the performance of back-testing for the pairs trading strategy when short selling was prohibited.

Table 3. The difference between the pairs trading and the pairs trading without short selling.

Symbol	S&P100	PT	PT w/o SS	Difference
02-Jan-01	1.000	1.000	1.000	0.00%
02-Jan-04	0.824	1.710	1.693	0.99%

03-Jan-07	0.990	2.353	2.686	14.15%
04-Jan-10	0.784	4.205	4.798	14.10%
02-Jan-13	0.995	5.713	7.124	24.70%
04-Jan-16	1.344	7.421	10.301	38.81%
02-Jan-19	1.676	10.140	14.607	44.05%
02-Jan-20	2.186	11.365	16.490	45.09%

Table 3 shows the difference between the general pairs trading strategy(PT as the abbreviation) and the pairs trading strategy when short selling is prohibited(PT w/o SS), and we can see that the pairs trading strategy without short selling earns 45.09% more than the overall performance of the pairs trading strategy with short selling. The difference came from the selection of stocks and income of general stock trades so that the general expectation of stock price is positive.

Table 4. The Sharpe ratio

	S&P100	PT	PTw/o SS
Expected Portfolio Return (Rx)	5%	14%	16%
Risk Free Rate (Rf)	3%	3%	3%
Standard Deviation of Rx	16%	6%	6%
SHARPE RATIO =	0.16	1.84	2.34

Table 4 shows the results of calculating the Sharpe ratio of the S&P100 index, PT and PT w/o SS strategies. The values of the Sharpe ratio were calculated as 0.16, 1.84, and 2.34, respectively. The Sharpe ratio in the PT strategy was measured as closer to 2, which shows that the strategy has considerable competitiveness. The annual expected portfolio return of PT is measured as 14%.

Table 5. Dependency of the performance with varying K. α =0.05, and ρ =100.

K	0.0001	0.0005	0.001	0.002	0.003	0.005
Performance	13.4244	12.1852	11.3795	10.4929	9.7384	8.7154

Finally, we vary one of parameters K, α, and ρ at a time and examine the dependence of the performance by the pairs trading strategy. Table 5 shows the change of performance when K is varying. When the K value was 0.0001 among the investigated values, the highest performance was obtained. If the K value was high or low, the performance was adversely affected. The larger the value of K, the larger the gap between k_1 and k_2. If the gap between k_1 and k_2 is too narrow, transactions are made frequently, but the profits from transactions are limited. On the other hand, if the gap is too wide, the trading opportunities are reduced, and performance may decrease.

Table 6. Dependency of the performance with varying α, and ρ. K=0.001.

$\rho \setminus \alpha$	0.1%	0.5%	1%	5%	10%	99%
2	1.54	2.84	3.04	2.93	3.13	3.52
5	3.39	5.97	6.81	7.98	7.57	7.56
10	3.61	5.36	6.61	8.55	8.85	9.39
20	3.43	5.65	7.72	9.26	10.79	11.38
50	3.38	5.64	7.64	9.32	10.67	11.42

100	2.99	5.88	8.37	10.46	12.06	12.99
500	3.49	8.16	10.87	13.01	12.77	14.49
1000	3.45	8.91	11.20	13.49	13.15	14.95

The Table 6 measures the sensitivity of performance: when K is fixed at 0.001, ρ is changed from 1 to 100, and α is changed from 0.001 to 0.99. Overall, as ρ increases and α increases, it can be seen that the value of performance increases. However, when α is greater than 0.005, the difference in performance is not noticeably large. On the other hand, the value of ρ seems to show the best performance values around 20.The most notable part is the performance value when α=0.99. The performance when α=0.99 means the value of performance when all companies handled in this research invest evenly in almost all pairs that can be constructed by all companies. This simply means that this pairs trading strategy does not require a cointegration test in the pair selection process. This is because buying and selling levels of this strategy are not determined by the difference between the price levels of the two stocks constructing a pair, but by the ratio of the price levels of two stocks.The effect of ρ on performance is even more dramatic. In our setting, the default value of ρ is 100, but if we set the value of ρ to a value greater than 100, as shown in the Table 2, we can expect better overall performance. According to Tie et al. (2018), the only limit for ρ is just $\rho > \max(\mu_1, \mu_2)$. They said "$\rho$ serves as a combined discounting and risk aversion rate." (p.659) Also, "Larger ρ encourages quicker profits, which leads to more buying and shorter holding." (p.673). That is, ρ penalizes late earnings. Therefore, ρ seems to be related to the length of the back-testing period as well. If the investment period is short, it would be more advantageous to set a large value of ρ to generate faster profits.

Figure 5. Changes of returns and equity curves of a pair trading when ρ changes.

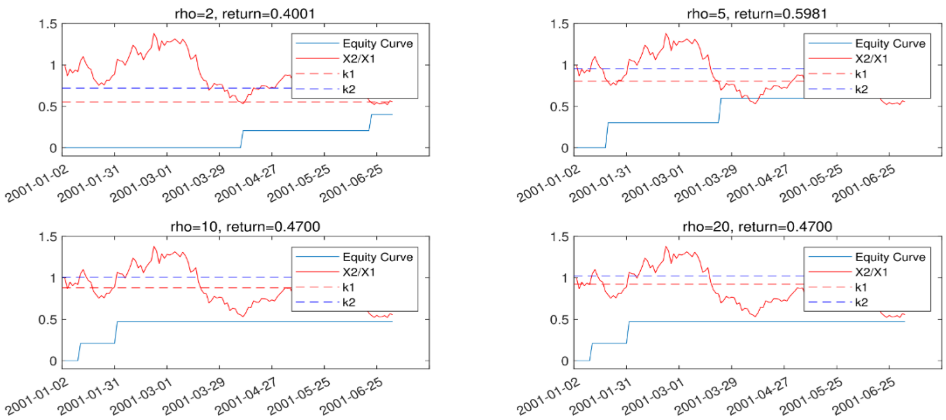
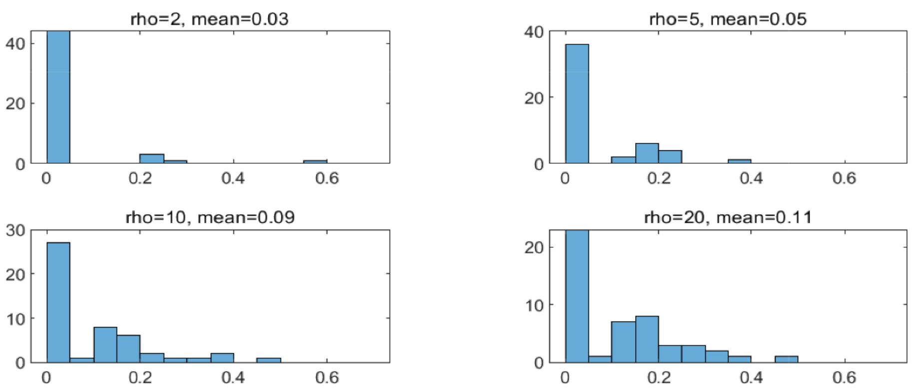


Figure 5 shows the changes in the equity curve expressed in Figure 3 when ρ is changed to 2, 5, 10, and 20. Although not all returns increased as rho increased, the increasein ρ moved k_1 and k_2, causing pairs trading profits to be realized faster, leading to an early rise in the equity curve.

Figure 6. Histograms of returns of pairs trading contracts when ρ changes.



Where α =0.01, the number of cointegrated pairs in a period 1 is 49 pairs. In Figure 6, the histograms show how the distribution of profits of 49 pairs during period 1 changes when ρ increases. As the ρ rises from 2 to 20, the number of pairs for which the profit during period 1 is 0 decreases, and the average of the profits rises from 0.03 to 0.11.

5. Conclusion

Through the research, we confirmed how much profit the pairs trading strategy of the stocks following Geometric Brownian Motion has in the market, using the data of the stocks constituting the S&P100. Pairs trading strategy has a distinctive advantage in that it can generate profits regardless of market conditions. If there is a way to generate optimal profits, it will be a very attractive strategy for the market.

For the optimal solution of the pairs trading strategy used in this study, we referred to Tie et al.(2018), and several conditions were reinterpreted to apply the research results to the actual market. We implemented a pairs trading strategy using 20-year data of the stocks included in the S&P100 portfolio, and as a result, the pairs trading strategy yielded more than three times the return in the S&P100 of the same period. The Sharp ratio of this pairs trading strategy is higher than 1.84, which can be classified as a good strategy for investment.In particular, it was confirmed that the ρ and α variables have significant influence on the performance of the portfolio, and that the choice of the ρ and α can result in better performance.

We also examined the effectiveness of the cointegration test to this pairs trading strategy, which is frequently used for traditional pairs selection in general pairs trading strategies. As the significance level α of the cointegration test is closer to 0, the difference between the stocks making up pairs is more likely to follow the mean-reversion model. In our research, despite α was sufficiently small, the overall performance did not increase. Rather, the closer the alpha is to 1, the slightly higher overall performance. In other words, this result shows that applying the cointegration test to this pairs selection has no effect in the optimal pairs trading strategy under Geometric Brownian Motion.This makes this pairs trading strategy be simpler in pairs selection perspectives.

Further studies to find a method of pairs selection, or an optimal value problem for ρ and α forthis pairs trading strategy will also be very interesting.

References

Dai, M., Zhang, Q., Zhu, Q.(2010). Trend following trading under a regime switching model. SIAM Journal on Financial Mathematics 1(1), 780–810

Fallahpour, S., Hakimian, H., Taheri, K., Ramezanifar, E.(2016).Pairs trading strategy optimization using the reinforcement learning method: a cointegration approach. Soft Computing 20(1), 5051–5066

Gatev, E., Goetzmann, W.N., Rouwenhorst, K.G.(2006).Pairs trading: Performance of a relative-value arbitrage rule. Yale ICF Working Paper No. 08-03 8(3), 47

Guo, X., Zhang, Q.(2005). Optimal selling rules in a regime switching model. IEEE Transactions on Automatic Control 50(9), 1450—-1455

Johansen, S.(1988). Statistical analysis of cointegration vectors. Journal of Economic Dynamics and Control 12(2), 231—-254

Lin, Y.X., McCrae, M., Gulati, C.(2006). Loss protection in pairs trading through minimum profit bounds: A cointegration approach. Journal of Applied Mathematics and Decision Sciences 2006(1), 1–14

Mashele, HP, Terblanche, SE and Venter, JH. (2013). Pairs trading on the Johannesburg Stock Exchange. Investment Analysts Journal, 42(78), 13-26

Mikkelsen, A.(2018).Pairs trading: the case of norwegian seafood companies. Applied Economics 50(3), 303–318

Mudchanatongsuk, S., Primbs, J.A., Wong, W.(2008). Optimal pairs trading: A stochastic control approach

Nath, P.(2003). High frequency pairs trading with U.S. treasury securities: risks and rewards for hedge funds

Rad, H., Low, R.K.Y., Faff, R.W.(2015). The profitability of pairs trading strategies: distance, cointegration and copulamethods. Quantitative Finance 16(10), 1541—1558

Song, Q., Zhang, Q.(2013). An optimal pairs-trading rule. Automatica 49(10), 3007–3014

Tie, J., Zhang, H., Zhang, Q.(2018). An optimal strategy for pairs trading under geometric Brownian motions. Journal of Optimal Theory and Application 179(2), 654–675

Tourin, A., Yan, R.(2013). Dynamic pairs trading using the stochastic control approach. Journal of Economic Dynamics & Control 37(10), 1972—-1981

Vidyamurthy, G.(2004). Pairs Trading: Quantitative Methods and Analysis. Wesley, Hoboken, New Jersey

Zhang, Q.(2001). Stock trading: an optimal selling rule. SIAM Journal on Control and Optimization 40(1), 64–87

[i]https://www.ishares.com/us/products/239723/ishares-sp-100-etf

[ii]https://finance.yahoo.com/