University of Warsaw
Faculty of Economic Sciences

Working Papers

No. 2/2022 (378)

THE PROFITABILITY OF PAIRS TRADING STRATEGIES ON HONG-KONG STOCK MARKET: DISTANCE, COINTEGRATION, AND CORRELATION METHODS

Baiquan Ma
Robert Ślepaczuk

Warsaw 2022

UNIVERSITY OF WARSAW
FACULTY OF ECONOMIC SCIENCES
WORKING PAPERS

The profitability of pairs trading strategies on Hong-Kong stock market: distance, cointegration, and correlation methods

Baiquan Ma[a], Robert Ślepaczuk[b]*

[a] *University of Warsaw, Faculty of Economic Sciences, Quantitative Finance Research Group*
[b] *University of Warsaw, Faculty of Economic Sciences, Quantitative Finance Research Group, Department of Quantitative Finance*
* *Corresponding author: rslepaczuk@wne.uw.edu.pl*

Abstract: This research aims to compare the profitability of correlation-based pair trading strategy, cointegration-based pair trading strategy, and distance-based pair trading strategy on the Hong Kong stock market. We try to build an effective pair trading strategy based on 50 stocks listed in the Hang Seng index joining them in market-neutral pairs. The dataset has a daily frequency and covers the period from 07/01/2013 to 07/01/2020. The result shows that all three methods are profitable in the Hong Kong stock market and can beat the market with regard to risk-adjusted return metrics. This result is quite sensitive to the varying number of pairs traded and rebalancing period and less sensitive to financial leverage degree. Moreover, the cointegration method is superior as compared to the correlation method and distance method.

Keywords: pairs trading strategy, market neutral strategy, algorithmic trading strategy, distance method, cointegration, and correlation trading, efficient market hypothesis, investment portfolio, algorithmic trading, Hang Seng index

JEL codes: C4, C14, C45, C53, C58, G13

Acknowledgements: This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors

Introduction

Pair trading is a type of statistical arbitrage strategy and a hedging strategy widely used in foreign securities markets. Looking back at the development of the pair trading strategy, the pair trading strategy originated in the 1920s when Wall Street trader Jesse Livermore adopted the Sister Stocks trading strategy in actual investment. In 1985, Nunzio Tartaglia, an astrophysicist at the Wall Street investment bank Morgan Stanley, formed a quantitative analysis team composed of famous physicists, computer scientists, and mathematicians. The team used mathematical models to calculate stock trading portfolios and used computer-automated trading programs to achieve immense success in actual investment. The trading portfolio strategy they used was the pair trading strategy. Different from traditional subjective analysis, Morgan Stanley's trading program adopts quantitative analysis methods for stock pair selection, trading parameter setting, and trading rules formulation and is automatically completed by computer programs. Since then, this quantitative investment strategy has been widely known and spread among traders. Today, pair trading is the basis of many models and trading rules used by various hedge funds and stock investors.

Pair trading is an excellent quantitative investment strategy as well as a relative value investment strategy. It selects long-term stable price differences and finds stocks that have been overvalued or undervalued recently, and obtains volatility returns from spread expansion and then convergence. The long-short, paired investment portfolio effectively avoids the uncertain market risk of the overall stock market in the future and can still obtain stable and considerable returns even during the period when the overall market is down. In recent years, the pair trading strategies have been applied to the Chinese stock market by a large number of quantitative investment practitioners.

As we all know, most of the time series of stock prices in stock trading decisions do not obey the stationary condition. Therefore, verifying whether there is a stable linear combination between variables that do not have a long-term stable and balanced relationship has become a hot research topic. Due to the characteristics of simultaneous short and long positions, a pair trading strategy can better avoid market risks, thus it becomes a favored strategy by investors in the quantitative investment field. China's research on pair trading started relatively late. Since the Chinese stock market officially launched the margin trading and securities lending business on March 31, 2010, the

Chinese stock market has had a short-selling mechanism since then. It makes the pair trading strategy becomes a viable practical method in the Chinese stock market. The Chinese scholars have begun to refer to foreign practices to study and improve the application of pair trading strategies in the Chinese market.

In that context, the paper compares the feasibility and effectiveness of the use of correlation-based pair trading strategy, cointegration-based pair trading strategy, and distance-based pair trading strategy in the Hong Kong stock market through the study of pair trading strategies. The paper addresses two main hypotheses: (RH1) *Is the pair trading strategy based on the distance, correlation, and cointegration method profitable in the Hong Kong stock market?* and (RH2) *Which method is a superior pair selection model for pair trading among three methods?* Additionally, based on these main research hypotheses, a few research questions are constructed: (RQ1) *Is the result obtained robust to the varying number of pairs selected?* (RQ2) *Is the result obtained robust to varying rebalancing periods?* (RQ3) *Is the result obtained robust to varying degree of financial leverage?*

In order to refer to the main hypotheses and answer the research questions mentioned above, empirical research is conducted based on 50 stocks listed in the Hang Seng index. The period of this dataset is ranging from 07/01/2013 to 07/01/2020. The data is collected daily. The pairs are formed from the pool of 50 stocks and then we try to apply those three different methods to select pairs. For examining distance, the process of price standardization is implemented. For examining correlation, the Pearson correlation coefficient is employed and for examining cointegration, the Engle-Granger two-step method is implemented. Furthermore, the technique to generate a buy/sell signal when we apply a pair trading strategy on each pair of stocks is the breakout volatility model (BVM). The upper and lower thresholds for BVM are calculated based on the exponential moving average and rolling standard deviation of the price spread for each pair. The performance of our strategy is evaluated based on six criteria: absolute return, annualized return, annualized standard deviation, maximum drawdown, Sharpe ratio, and information ratio.

The paper is structured as follows, the first section is an overview of literature about correlation, cointegration, and distance methods. The second section describes all the details concerning the data set. The third section explains the methodology of how the pair trading strategy is constructed and the technique to generate buy/sell signals for such a strategy. The fourth section discusses the empirical results. The fifth section conducts some sensitivity analysis to determine whether the result is robust to changes in initial assumptions. The sixth section draws conclusions.

1 Literature review

1.1 Correlation

Correlation, in the finance and investment industries, is a statistic that measures the degree to which two securities move in relation to each other. Correlations are used in a wide range of investment strategies, one of them is portfolio management, computed as the correlation coefficient, which has a value that must fall between -1.0 and +1.0. Correlation shows the strength of a relationship between two variables and is expressed numerically by the correlation coefficient. A perfect positive correlation means that the correlation coefficient is exactly 1. This implies that as one security moves, either up or down, the other security moves in lockstep, in the same direction. A perfect negative correlation means that two assets move in opposite directions, while a zero correlation implies no linear relationship at all.

One advantage of the correlation coefficient is that it is easy to use. The correlation technique is quite simple. All standard software packages can calculate the correlation coefficient. Thus, Correlations are also used in pairs trading. Wang et. al (2009) analyzed the relative performance of different correlation measures of high-frequency pairs trading by backtesting three diverse types of measures over as many pairs as possible. Their most important conclusion is that different statistical correlation measures do show significant differences in terms of risk and return. Hauke and Kossowski (2011) pointed that the Pearson coefficient is particularly useful when the relationship between the two variables is described by a monotonous function and does not assume any particular distribution of the variables. Chen et. al (2012) suggest monitoring the co-movement of stock pairs by computing the Pearson correlation coefficient. High values indicate the most suitable pairs to trade in the future. Ramos et. al (2017) pointed out an issue about correlation-based pairs trading that is the data frequency on which to use correlation measures. They pointed that correlation is intrinsically a short-run measure because it is based on returns, which is a short memory process. This fact implies that the higher the trading frequency, the more likely a correlation-based pairs trading strategy will work and thus the more potential for profits.

1.2 Cointegration

Engle and Granger (1987) proposed the cointegration theory. The cointegration approach is an econometric technique for testing the relationship between non-stationary time series variables. If two or more series have a unit root, that is I(1), but a linear combination of them is stationary, I(0), then the series are said to be cointegrated.

There are many studies on pairs trading based on the cointegration method in the international stock market. Lin et. al (2007) applied the cointegration method to the Australian stock market. They added a minimum profit constraint and used two Australian bank stocks for research. Puspaningrum et. al (2010) proposed an assumption that when the residual items of the cointegration test obey the AR(1) process, how to set the paired transaction trigger points and optimal boundaries. They also propose a numerical algorithm for estimating the trading range and average trading period. Caldeira and Moura (2013) use a cointegration-based trading strategy on the Sao Paolo exchange. They find that the strategy generates a 16.4% excess return per annum with a Sharpe ratio of 1.3 from 2005 to 2012. Afawubo (2015) used the S&P500 index constituents to conduct research and chose different methods of stock pairs selection. He shows that after controlling costs and risks, using the distance method to obtain extremely low excess returns, and the cointegration method can provide stable and reliable returns. Bui and Ślepaczuk (2020) study the performance of three different pair trading strategies. Generalized Hurst Exponent, Correlation, and Cointegration are implemented and tested on the 103 stocks listed in NASDAQ-100 index from 2000 to 2018. The study concludes that the results of all three trading strategies are quite sensitive to varying number of pairs traded and rebalancing period and much less sensitive to financial leverage degree. Moreover, the Hurst method is better than the cointegration method but is not superior as compared to the correlation method. Liew and Wu (2013) believe that the asset returns do not follow a normal distribution, thus the linear assumptions implicit in the minimum distance method and cointegration method are not valid. They get the conclusion that the copulas method is more suitable for simulating the linear relationship between two time series. Rad et. al (2016) further examine the profitability of pairs trading strategies based on distance, cointegration, and copula methods. They find that all strategies show positive and significant alphas after accounting for various risk factors. In addition, all strategies perform better during periods of significant volatility, and the cointegration method is the superior strategy during turbulent market conditions.

1.3 The Distance Method

The distance technique was developed and revised by Gatev et. al (2006). Their study is performed on all liquid US stocks from the CRSP daily files from 1962 to 2002. they set the formation period of 12 months, the trading period of 6 months, and a fixed trading threshold in the pairs trading strategy. Their research found that the distance method pairing trading strategy achieved extremely high excess returns, that strategy has a higher Jensen alpha value, lower risk, and reasonable transaction costs after risk adjustment. They also found that the profitability of pairs trading strategies declined over time Their work is considered by many authors as the best-known work devoted to pairs trading. Do and Faff (2010, 2012) further examine the distance method strategy of Gatev et. al (2006) to investigate the source of its profits and the effects of trading costs on its profitability using CRSP data from 1962 to 2009. They confirm declining profitability in pairs trading, due to an increasing share of nonconverging pairs. With the inclusion of trading costs, pairs trading according to Gatev's baseline methodology becomes largely unprofitable.

Huck (2013) studied the sensitivity of pairs trading strategy parameters based on the minimum distance method. He found that the minimum distance method is extremely sensitive to changes in the length of the formation period, so a reasonable adjustment of trading parameters may produce excess returns. Smith and Xu (2017) studied the methods of selecting stock pairs in the pair trading strategy from 1980 to 2014. Their research considers two methods: the minimum distance method and the cointegration method, and the trading parameters involved in the trading system. The empirical results show that the trading parameters in the trading system are related to the profitability of pairs trading. In addition, they found that the cointegration method only produced significant gains in the 1980s. The distance method performs well, However, if the transaction costs are considered, both of the methods basically impossible to obtain a stable rate of return. Rad et. al (2016) used 50 years of long-term comprehensive data to evaluate distance, cointegration, and copula methods. They found that the performance of the copula method is weaker than the distance method and the cointegration method in terms of excess returns and various risk adjustment indicators.

Fernando et. al (2017) proposed a pairs trading method of Archimedes copulas and applied it to the constituent stocks of the S&P500 index from 1990 to 2015. The empirical results show that although the pairs trading strategy based on the minimum distance has greater volatility than the copula-based pairs trading strategy, it can obtain more trading opportunities. When paired stocks are under different weight structures, trading imbalance occurs, the copula-based pair trading strategy produces higher risk-adjusted excess returns than the distance-based pair trading strategy, which reduces the trading risk. Liu et. al (2016) selected the data of the oil companies listed on the New York Stock Exchange in 2008 and the 5-minute interval from June 2013 to April 2015 as the research objects. They used the double mean response process to Model the mispricing in a more dynamic way, then they compared it with the minimum distance method and the cointegration method, the result shows that the pairs trading strategy of the new model obtained better returns. Bowen et. al (2010) used the minimum distance method to analyze the 60-minute data of FTSE100 index stocks from January 2007 to December 2009. They found that the return of a pair's trading strategy is very sensitive to transaction costs and execution speed. When the transaction costs are increased by 15 points, the pairs trading strategy cannot make a profit.

2 Data

We selected our data set from the Hang Seng index constituents listed in the document titled: *Hang Seng Indexes Announces Index Review Results* which was released usually in May from 2013 to 2020 by Hang Seng Indexes Company Limited. [1]We also selected the Hang Seng index itself which was used as the benchmark for our strategy. After the data processing and collection, there are 50 stocks included in our data set. The data frequency was daily. The data onto each stock was carefully investigated by graph and cleaned for outliers if any. Basic statistics of all the stocks were also gathered for

[1] Available at: https://www.hsi.com.hk/eng/newsroom/press-releases

cleaning data purposes. According to our trading strategy, the entire pair trading process is divided into a matching phase and a trading phase. The literature generally sets the matching period as 1 year and the trading period as 6 months. In this study, we use the same matching period as 1 year, but the trading period was also as 1 year. The research period was from 07/01/2013 to 07/01/2020. As shown in Table 1

Table 1. Summary of trading intervals

Matching period	**Trading period**
2012-07-01~2013-07-01	2013-07-01~2014-07-01
2013-07-01~2014~07-01	2014-07-01~2015-07-01
2014-07-01~2015-07-01	2015-07-01~2016-07-01
2015-07-01~2016-07-01	2016-07-01~2017-07-01
2016-07-01~2017-07-01	2017-07-01~2018-07-01
2017-07-01~2018-07-01	2018-07-01~2019-07-01
2018-07-01~2019-07-01	2019-07-01~2020-07-01

Note: The trading intervals used for the purpose of this study. The matching phase and a trading phase lasted 1 year.

3 Methodology

3.1 Pair selection based on Correlation

In this work, we will use the Spearman correlation coefficient to gain the correlation for the two series of log returns. The Spearman correlation coefficient for a sample A_i, B_i of size n can be described as follows: first, consider the ranks of the samples rg_{Ai}, rg_{Bi}, then the Spearman correlation coefficient r_s is calculated as:

$$r_s = \rho_{rg_A, rg_B} = \frac{cov(rg_{A,} rg_B)}{\sigma g_A * \sigma g_B} \tag{1.}$$

where ρ denotes the Pearson correlation coefficient, applied to the rank variables, $cov(rg_{A,} rg_B)$ is the covariance of the rank variables, σg_A and σg_B are the standard deviations of the rank variables.

Given paired data {(A1, B1), …, (An, Bn)} consisting of n pairs, we can rewrite the formula for r_s by replacing the estimates of sample covariances and variances as below:

$$r_s = \frac{\sum_{i=1}^{n}(A_i - \bar{A}) + (B_i - \bar{B})}{\sqrt{\sum_{i=1}^{n}(A_i - \bar{A})^2 \sum_{i=1}^{n}(B_i - \bar{B})^2}} \tag{2.}$$

where $\bar{A}$ and $\bar{B}$ are the mean of paired data (A, B).

In this context, we would like to use the correlation formula for the sample. we will conduct a calculation of correlation based on the data of the past one year at the beginning of each 1 year during the period 2013 – 2020. The 10 pairs that have the highest correlation would be used in pair trading strategy during that 1 year.

3.2 Pair selection based on Cointegration

We employ the Engle-Granger two-step approach to test for cointegration. Usually, on the first step, we check whether two time series are integrated with the same order d using the Augmented Dickey-Fuller test. The basic idea of the Engle-Granger two-step test method is assuming that two time series have a cointegration relationship, then there is a stable equilibrium relationship between these two time series, their specific composition of linear combination sequence is also stationary. Therefore, by using the Augmented Dickey-Fuller test to evaluate whether the residual sequence of the regression equation is stationary, we can know whether there is a cointegration relationship between the two time series. Generally speaking, in order to simplify the testing process, most scholars directly use standard OLS regression and testing whether residuals obtained in the regression are stationary using the ADF test to determine whether the two time series have a cointegration relationship. According to the research of Bui and Ślepaczuk (2020), for the second step, we use the KPSS test instead of the ADF test. The pair selection is conducted every 1 year and 10 pairs with the lowest KPSS test statistics, which indicates a higher chance of cointegration between two time series, would be used to trade during that 1year. The Engle-Granger approach is as follows:

Step 1: check whether two time series are integrated with the same order d using the Augmented Dickey-Fuller test. For stock prices, they are normally integrated at order 1.

$$X, Y \sim I(d) \tag{3.}$$

Step2: Estimation of the cointegrating vector using standard OLS regression and testing whether residuals obtained in the regression are stationary using the KPSS test.

$$Y_t = \alpha + \beta X_t + \varepsilon_t \tag{4.}$$

$$\varepsilon_t = Y_t - \alpha - \beta X_t \tag{5.}$$

$$Check\ \varepsilon_t \sim I(0)$$

KPSS stationary test is quite similar to the ADF test but more powerful. The most crucial difference between KPSS and ADF test stays in the null hypothesis. For the KPSS test, the null hypothesis is that the series is stationary, and the alternative hypothesis is that the series is non-stationary. Hence, it is an upper-tail test, and we reject the null when our test statistic is higher than the critical value, which is normally 0.463 for a 5% significance level.

3.3 Pair selection based on the Distance Method

Similar to Correlation and cointegration, we select pairs based on the distance method between 2 time series of prices. The distance between each pair of assets s is determined by:

$$D = \sum_{i=1}^{n} \sqrt{(P_{Ai} - P_{bi})^2} \tag{6.}$$

where P_{Ai} and P_{bi} are the standardized price of assets A and B at moment i. D is the distance between both assets. The standardized price of an asset is determined by:

$$P_{it} = \frac{P_t - \overline{P_i}}{\sigma_i} \tag{7.}$$

where P_{it} is the standardized price of asset *i* at moment t, P_t is the price of the asset at moment t, $\overline{P_i}$ is the mean value of asset I, and σ_i is the standard deviation of asset i.

The pair selection is conducted every 1 year, and 10 pairs were that had the least distance between them, which indicates that the more consistent the price trend of stocks is, the more suitable it is as a stock pair.

3.4 Trading strategies description

For each possible pair, we calculate the spread according to the formula:

$$spread = \ln(P_A) - h * \ln(P_B) \tag{8.}$$

where P_A and P_B are the prices of stock *A* and *B*, *h* is the weight factor.

There are many ways to calculate the weight factor h. For example, we can use the equal weight method. In this case, h = 1 and this is the way used in most of the literature. In this case, the position in the pair is market neutral. This method was used by Gatev et. al (2006), and since then, it has become the most popular procedure to fix h. In this study, we use the Volatility method to calculate h, it is based on the idea that both stocks are normalized if they have the same volatility. This approach was used by Ramos et. al (2017), and Bui and Ślepaczuk (2020). Then, the weight factor h can be calculated as:

$$h = \frac{std(\ln(ret_A))}{std(\ln(ret_B))} \tag{9.}$$

where *std* is the standard deviation ret_A and ret_B are the log-returns of stock A and B.

When we buy the pair, it means that for each share of A that we buy, we short sell h shares of B. According to the above formula to estimate h, we will have the same volatility for both the position in A and in B.

After calculating the spread, we use breakout volatility for each series of spreads to generate a signal to go long or short the pair. The upper and lower threshold is calculated as follow:

$$upper\ threshold = EMA_n(spread) + m * std_k(spread) \tag{10.}$$

$$lower\ threshold = EMA_n(spread) - m * std_k(spread) \tag{11.}$$

where EMA_n is the rolling Exponential Moving Averages of the spread of each pair with certain window size n ranging from 10 to 180 (days), std_k is the rolling standard deviation of the spread with certain window size k ranging from 10 to 120 (days), multiplier m takes a certain value ranging from 0.5 to 3. These parameters will be optimized during our study. we will use the data available up to the moment of selection to optimize parameters.

For pair trading strategy, we will short the recent winners (the price was increasing before) and long the recent losers (the price was decreasing before) as we believe that their prices will behave according to the mean-reverting pattern. It means that the current high price will go down and the current low price will go up in the future. In general, the rule for the entry and exit is as follow:

(a) if our position is being flat at time t-1, we will short the pair if the spread reaches the upper threshold or long the pair if the spread reaches a lower threshold at time t.

(b) if our position is being long at time t – 1, we only switch to the short position if the spread reaches the upper threshold at time t.

(c) if our position is being short at time t – 1, we only switch to the long position if the spread reaches the lower threshold at time t. This trading strategy refers to Bui and Ślepaczuk (2021).

The initial investment is assumed to be 10000 HK$ at the beginning of the rebalancing period. This investment is divided equally for *N* pairs, so each pair takes up HK$10000/N. In our case, we have 10 pairs, so the amount of 1000 HK$ is invested in each pair. Due to the rule in the Hong Kong stock market, when we buy a stock, we have to pay the Handling Fee, the Securities Management Fee, the stamp duty, and so on. In that case, our total transaction percentage fee is assumed to be 0.02% of the value of the pair. So, the transaction cost for trading 1 pair is 0.02% * (P_A + h*P_B). Moreover, we can also take advantage of financial leverage to improve the returns of the pair trading strategy. So, instead of investing $10 000/N, we can invest $10 000/(N/2) or $10 000/(N/3) in each pair by borrowing money. The financial leverage can help us magnify the positive returns, but it also magnifies the negative returns if we make a loss. For the benchmark, we just use the Buy&Hold strategy on the Hang Seng index in the period 2013 – 2020 and compare our strategy's performance with this benchmark.

Table 2. The list of initial assumptions and optimized parameters

Parameters	Assumptions/parameters
Window size of EMA	To be optimized: {10, 20, 45, 60, 90, 120, 150, 180}
Window size of rolling std	To be optimized: {10, 20, 45, 60, 90, 120}
Multiplier m	To be optimized: {0.5, 1, 1.5, 2, 2.5, 3}
Rebalancing period	1 year
Number of pairs	10 pairs
Initial investment	HK$ 10 000
Degree of Financial Leverage	100% (investment in each pair = Initial investment/N with N = no. of pairs)
Spread of each pair	ln(P_A) – h*ln(P_B)
Transaction cost	0.02% * (P_A + h * P_B) for trading 1 pair

Note: All optimized parameters were set on in-sample window.

Before investigating the result, we would like to describe the list of assumptions made in our methodology section. Table 2 gives us a clear picture of the factors that may have an impact on our result and based on that we can then evaluate the sensitivity of our results against the change of these factors.

3.5Performance metrics

The following measures were used to provide descriptive statistics for the data and further these measures were applied in the evaluation of strategies' efficiency. These metrics were based on Kość et al. (2019) and Kijewski and Ślepaczuk (2020).

The Absolute return is expressed as a percentage (%) and calculated as:

$$Absolute\ return = P = \frac{P_{start} - P_{end}}{P_{start}} \quad (12.)$$

where P_{start} is the total value of stocks and cash held at the beginning of a portfolio transaction, P_{end} is the total value of stocks and cash held at the end of a portfolio transaction.

The Annualized return (ARC) is expressed as a percentage (%) and calculated as:

$$ARC = (1 + P)^{\frac{252}{n}} - 1 \quad (13.)$$

where P is the Absolute return, n is the sample size.

The Annualized standard deviation is expressed as a percentage (%) and calculated as:

$$ASD = \sqrt{252} * \sqrt{var(R)} \quad (14.)$$

where R is the percentage rate of return.

The Sharpe Ratio is calculated as:

$$SR = \frac{ARC - r_f}{ASD} \quad (15.)$$

where r_f is the risk-free rate.[2]

The Maximum Drawdown (MD) is expressed as a percentage (%) and calculated as:

$$MD = max_{(x,y)\in\{(t_1,t_2)^2 : x \le y\}} \frac{P_x - P_y}{P_X} \quad (16.)$$

where the P_t is the price level.

[2] The value of risk-free rate was set as 1.49% in the formula (15). It was calculated by the yearly HIBOR rate from 2013 to 2020. It is available at: https://www.hkab.org.hk/DisplayInterestSettlementRatesAction.do

The information ratio (IR) is simply the ratio of annualized return (ARC) and annualized standard deviation:

$$IR = \frac{ARC}{ASD} \tag{17.}$$

4. Empirical Results and Discussion

The performance of each method for pair trading strategy and benchmark are presented in Table 3. Looking at the result, we can see that the cointegration method perform the best among the three tested methods, and even better than the benchmark strategy. It generates the highest Absolute return (AR) with 20.95%, the highest Annualized Return (ARC) with 2.81%, the highest Sharpe ratio (SP) with 0.439 and the highest information ratio (IR) with 0.934. All of these are better than the benchmark strategy generates. Although the Annualized Standard Deviation (ASD) with 3.01% and the Maximum Drawdown (MD) with 4.11% for the cointegration method are not the best, they are still better than for benchmark strategy characterized by ASD on the level of 17.89% and MD equal to 35.67%. The correlation method generates the lowest ASD with 1.71%, and MD equal to 4.11% and IR equal to 0.830 which is better than for the benchmark strategy. AR for the correlation method equal to 10.13% and ARC equal to 1.42% are lower than the benchmark strategy's AR equal to 17.89% and ARC equal to 2.07%. The distance method does not perform well, although it generates the lowest MD with 4.03%. Its AR, ARC, SR, and IR are the worst among 3 methods. All three tested methods generate a much lower ASD and MD compared to the benchmark.

Table 3. Performance of all methods and benchmark

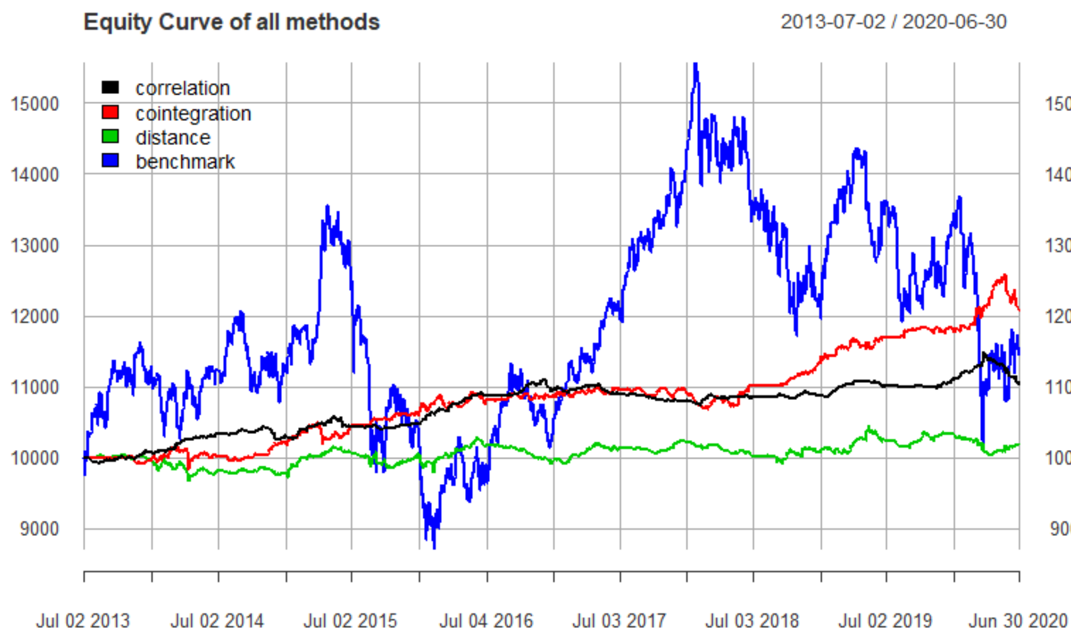
Method	AR (%)	ASD (%)	ARC (%)	SP	MD (%)	IR
Correlation	10.13	**1.71**	1.42	-0.043	4.12	0.830
Cointegration	**20.95**	3.01	**2.81**	**0.439**	4.11	**0.934**
Distance	2.10	2.69	0.30	-0.442	**4.03**	0.113
Benchmark	15.05	17.89	2.07	0.032	35.67	0.115

Note: AR: Absolute return; ASD: Annualized Standard Deviation; ARC: Annualized Return; SP: Sharpe Ratio; MD: Maximum Drawdown; IR = ARC/ASD: Information Ratio. Bold font indicates the best results in case of each performance statistics.

In general, the results implies that the three methods applied in pair trading strategy are profitable in Hong Kong stock market. In addition, the pair trading strategy based on the cointegration method significantly outperform the benchmark strategy. Moreover, comparing the performance of three tested methods, we can see that the correlation method cannot outperform the cointegration method but performs better than distance and benchmark methods. The equity curves of all 3 methods and benchmark are illustrated in Figure 1.

Figure 1 gives us more information on how each strategy performs in different market conditions. We can see that in the bear market condition (market downturn), the benchmark performs worse than all methods, while in the bull market condition, the benchmark performs the best, while correlation and cointegration methods perform better than the distance method though still lower than the benchmark. Besides, we can see that during the whole rebalancing period, pair trading strategies in general generate much lower volatility of returns compared to the benchmark.

Figure 1. Equity curve of all methods and benchmark



Note: The result of pair trading strategy with signals based on correlation, cointegration and distance methods compared with the results of Buy&Hold strategy for Hang Seng Index. The trading period started on 07/01/2013 and ended on 07/01/2020. The number of pairs traded was 10. The rebalancing period was 1 year. Degree of financial leverage = 100%.

5. Sensitive Analysis

In this section, we would verify how the result we obtained above is robust to the varying number of pairs traded, varying rebalancing periods, and varying degree of financial leverage. In the result above, we trade 10 pairs during the period of 1 year with the initial investment of HSK $10000. To check the sensitivity of this result, we change the number of pairs traded to 5 and 20 pairs, the rebalancing period to 6 months and the financial leverage degree to 200% and 300% in each pair.

5.1 Varying number of pairs traded

Table 4 presents the performance of all three methods with 5 pairs, 10 pairs, and 20 pairs. As we can see in Table 4, with 5 pairs, the performance of all three methods is worse than before (with 10 pairs), The distance method even lead to negative profit with AR of -0.94%; ARC of -0.14%; SP of -0.416; and IR of -0.150, while cointegration method still performs the best among 3 methods. Although its AR of 12.13% and the ARC of 1.68% is now lower than the benchmark strategy. It still can beat the market with SP of 0.048 and IR of 0.416. The correlation method performs similarly but a little lower than before. It still outperforms the distance method. With 20 pairs, the performance of the correlation and the cointegration methods are still worse than the performance with 10 pairs but still better than the benchmark with IR of 0.217 and 0.797 respectively. We can observe that the performance of the distance method improves more compared to other number of pairs. It works well with IR of 0.688, which outperforms the correlation method and the market but is still lower than the cointegration method.

Table 4. Performance of all methods and benchmark for 5, 10 and 20 pairs

Method	No. of pairs	AR (%)	ASD (%)	ARC (%)	SP	MD (%)	IR
Correlation	5	8.70	1.83	1.22	-0.145	3.84	0.668
Cointegration	5	12.13	4.05	1.68	0.048	6.28	0.416
Distance	5	-0.94	3.91	-0.14	-0.416	9.12	-0.035
Correlation	10	10.13	**1.71**	1.42	-0.043	4.12	0.830
Cointegration	10	**20.95**	3.01	**2.81**	**0.439**	4.11	**0.934**
Distance	10	2.10	2.69	0.30	-0.442	4.03	0.113
Correlation	20	2.84	1.88	0.41	-0.575	4.29	0.217
Cointegration	20	12.17	2.12	1.69	0.094	4.24	0.797
Distance	20	9.53	1.94	1.34	-0.079	**2.12**	0.688
Benchmark	-	15.05	17.89	2.07	0.032	35.67	0.115

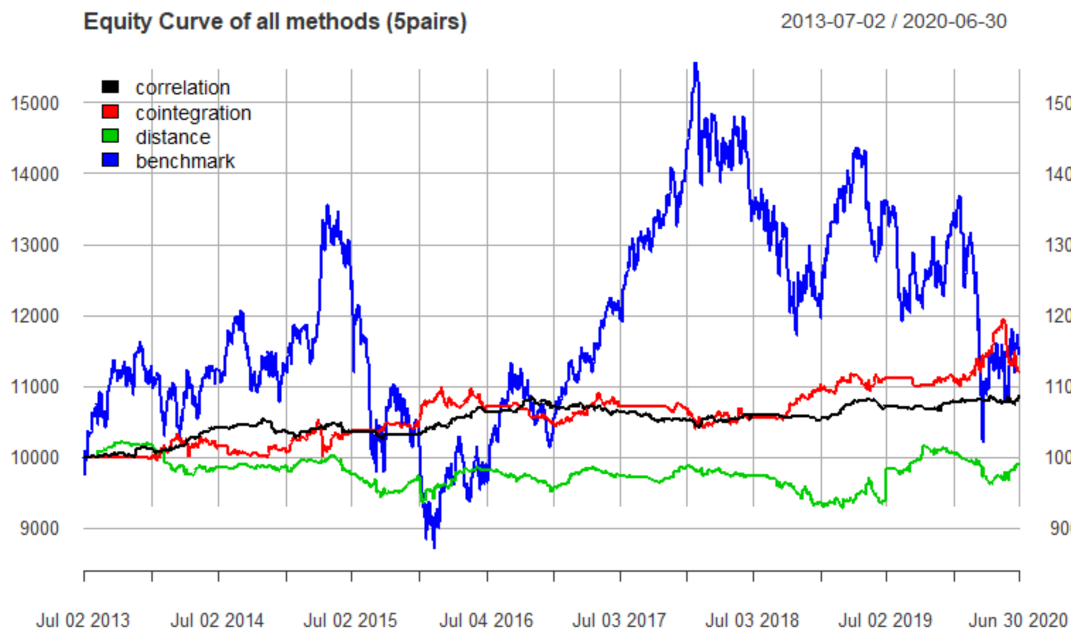
Note: AR: Absolute return; ASD: Annualized Standard Deviation; ARC: Annualized Return; SP: Sharpe Ratio; MD: Maximum Drawdown; IR = ARC/ASD: Information Ratio. Bold font indicates the best results in case of each performance statistics.

Based on these results, we can conclude that all three methods seem to be sensitive to the number of pairs traded. The performance of distance methods can be improved with the greater number of pairs. We can see that the distance method works well with 20 pairs and is quite poor with 5 and 10 pairs. However, it is not clear whether the performance of correlation and cointegration methods improves with the greater number of pairs. As we can see, the correlation and the cointegration methods perform better with 10 pairs and worse with 5 and 20 pairs.

The result confirms that pair trading strategy performs well and better than the market in the period of market crash but not only, and confirms the market-neutral characteristics of pair trading strategy.

The equity curves of all methods and the benchmark with 5 and 20 pairs are plotted in Figures 2 and 3 respectively.

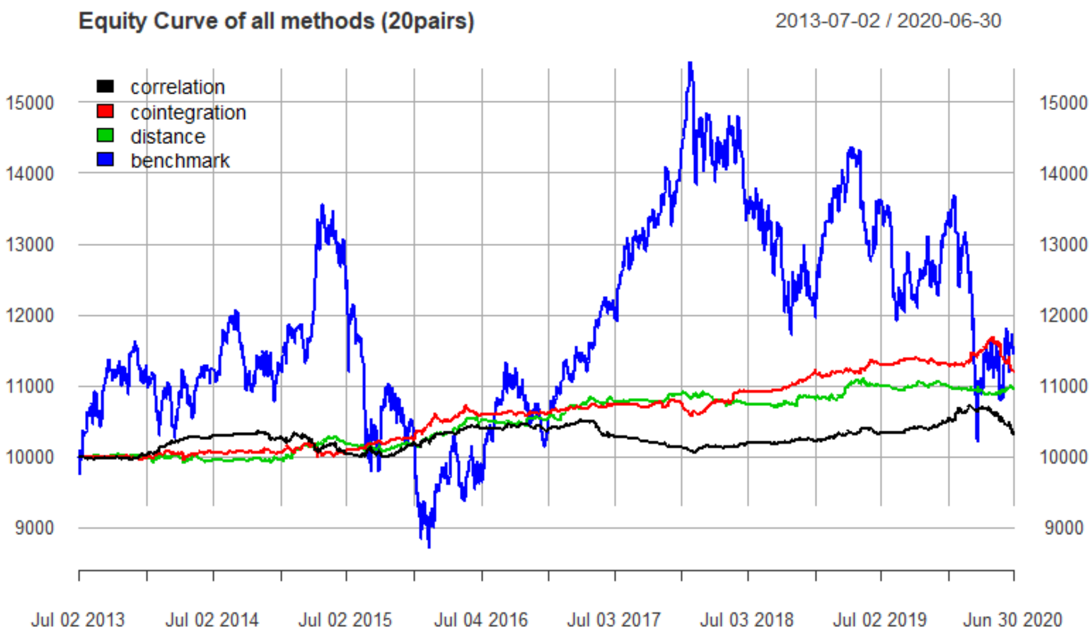
Figure 2. Equity curve of all methods and benchmark - 5 pairs



Note: The result of pair trading strategy with signals based on correlation, cointegration and distance methods compared with the results of Buy&Hold strategy for Hang Seng Index. The trading period started on 07/01/2013 and ended on 07/01/2020. The number of pairs traded was 5. The rebalancing period was 1 year. Degree of financial leverage = 100%.

In general, they all show that pair trading strategies with all methods (correlation, cointegration, and distance) work roughly the same or higher compared to the benchmark in the market downturn. Besides, the cointegration method's performance seems to be outstanding compared to the correlation and the distance method for any number of pairs. With 5 pairs, the distance method seems to perform poorer than the other two methods. With 20 pairs, the distance method seems to perform better than the correlation method but still poorer than the cointegration method.

Figure 3. Equity curve of all methods and benchmark - 20 pairs



Note: The result of pair trading strategy with signals based on correlation, cointegration and distance methods compared with the results of Buy&Hold strategy for Hang Seng Index. The trading period started on 07/01/2013 and ended on 07/01/2020. The number of pairs traded was 20. The rebalancing period was 1 year. Degree of financial leverage = 100%.

5.2 Varying rebalancing period

Table 5 shows the performance of all three methods with a rebalancing period of 6 months and 1 year (the original rebalancing period), and the number of pairs traded is fixed at 10 pairs for all 3 methods. As we can see in Table 5, with 6 months rebalancing period, the distance method generates the highest AR and ARC with 14.21% and 1.95% but it also generates the highest ASD with 2.65% among the 3 tested methods. Thus, the correlation method has the highest performance based on IR = 0.794, while the distance method has the highest SP = 0.175, and the cointegration method has lowest performance with SP = 0.048 and IR = 0.588. Hence, we can see that all these methods are profitable in the Hong Kong stock market and can beat the market (with SP = 0.032 and IR = 0.115) with 6 months rebalancing period.

Based on these results, we can observe that all three methods are quite sensitive to different rebalancing periods, but the correlation method is more stable than the other two as it keeps earning positive profit with different rebalancing periods, The performance of the cointegration method drops significantly with 6 months rebalancing period as compared to 1 year period. For the distance method, the 6-month rebalancing period earns much better positive profit than the 1-year period.

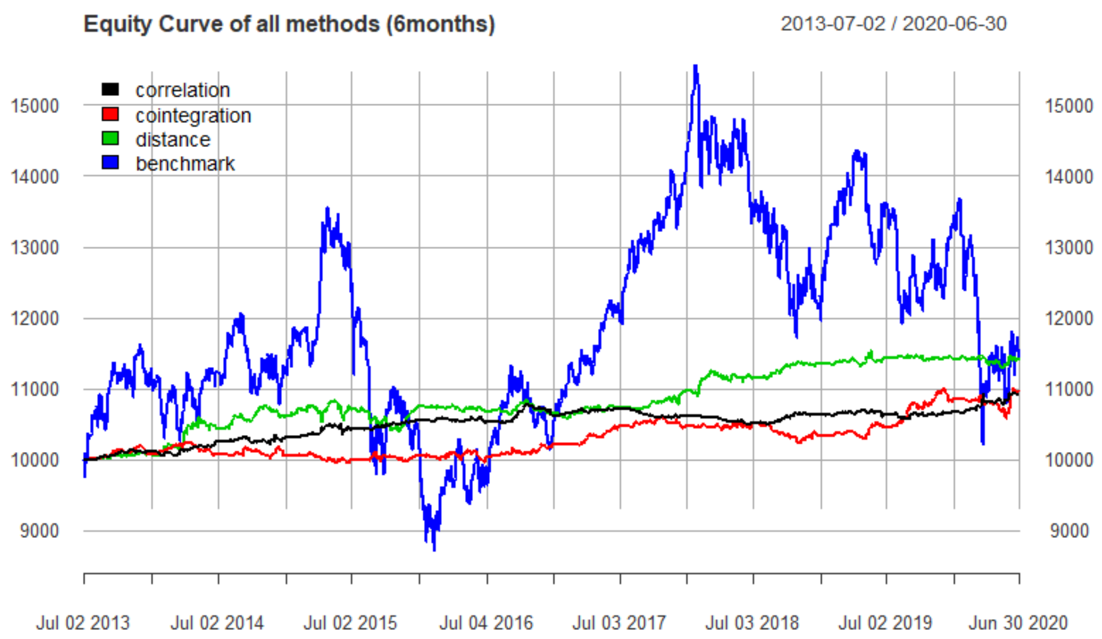
Table 5. Performance of all methods and benchmark - 6 months and 1 year

Method	Rebalancing period	AR (%)	ASD (%)	ARC (%)	SP	MD (%)	IR
Correlation	6 months	9.10	**1.61**	1.28	0.132	**2.65**	0.794
Cointegration	6 months	9.84	2.34	1.38	0.048	3.87	0.588
Distance	6 months	14.21	2.65	1.95	0.175	4.10	0.737
Correlation	1 year	10.13	1.71	1.42	0.043	4.12	0.830
Cointegration	1 year	**20.95**	3.01	**2.81**	0.439	4.11	**0.934**
Distance	1 year	2.10	2.69	0.30	**0.442**	4.03	0.113
benchmark	-	15.05	17.89	2.07	0.032	35.67	0.115

Note: AR: Absolute return; ASD: Annualized Standard Deviation; ARC: Annualized Return; SP: Sharpe Ratio; MD: Maximum Drawdown; IR = ARC/ASD: Information Ratio; Bold font indicates the best results in case of each performance statistics.

Figure 4 illustrates the equity curves of all strategies and the benchmark with 6 months rebalancing period.

Figure 4. Equity curve of all methods and benchmark - 6 months



Note: The result of pair trading strategy with signals based on correlation, cointegration and distance methods compared with the results of Buy&Hold strategy for Hang Seng Index. The trading period started on 07/01/2013 and ended on 07/01/2020. The number of pairs traded was 10. The rebalancing period was 6 months. Degree of financial leverage = 100%.

We can see all three methods outperform the benchmark in bear market conditions, but not only. In the bull market condition, no method can beat the market.

However, in general, the distance method seems to perform better than the other two methods in all market conditions.

5.3 Varying degree of financial leverage

Table 6 shows the result for varying degree of financial leverage, so instead of the investment of HK$10000/N in each pair, we try to invest HK$10000/(N/2) and HK$10000/(N/3) in each pair, considering that we can earn a higher return with a reasonable level of volatility. The number of pairs is fixed at 10 and the rebalancing period is fixed at 1 year. With the degree of Financial Leverage at 200% (HK$10000/(N/2)), the SP of correlation and cointegration methods improves with 0.353 and 0.726 respectively, which can beat the market with SP = 0.032. The reason for this is that with a double degree of financial leverage, we gain almost double AR and ARC for correlation and cointegration methods (20.03% and 2.70% for Correlation, and 48.13% and 5.90% for Cointegration) which was much better than the benchmark strategy. Though we also get almost double ASD (3.42% for Correlation and 6.03% for Cointegration), in total it enables our SR to improve as well. In terms of IR, we get almost double ARC but also double ASD and MD for each method, overall making IR better only for the correlation and the distance methods. The results are similar with the degree of Financial Leverage at 100%, only distance method cannot beat the market with IR = 0.098, both correlation and cointegration method with IR of 0.788 and 0.972 can beat the market. The cointegration method performs the best.

With the degree of Financial Leverage at 300% (HK$10000/(N/3)), The cointegration method performs much better than the benchmark strategy as it earns more than 4 times ARC (9.17%) compared to the market. Information Ratio (IR) of the correlation method can outperform the market with 0.763, which is mainly because the correlation method can earn higher ARC with lower ASD compared to the market. For the distance method, the IR of 0.083 still cannot beat the market and is also lower than the original performance (without financial leverage) as it earns less than 3 times ARC (0.66%) but also has almost 3 times ASD (7.95%). We can see the cointegration method still performs the best.

From the analysis of the result in Table 6, we can see that the return of all methods improves significantly with a higher degree of financial leverage, but it also results in higher standard deviation and maximum drawdown. Overall, only the distance method cannot outperform the market while the correlation and the cointegration methods perform well. Hence, we can say that the result we gain is less sensitive to the degree of financial leverage than to the number of pairs or rebalancing period as we discussed above.

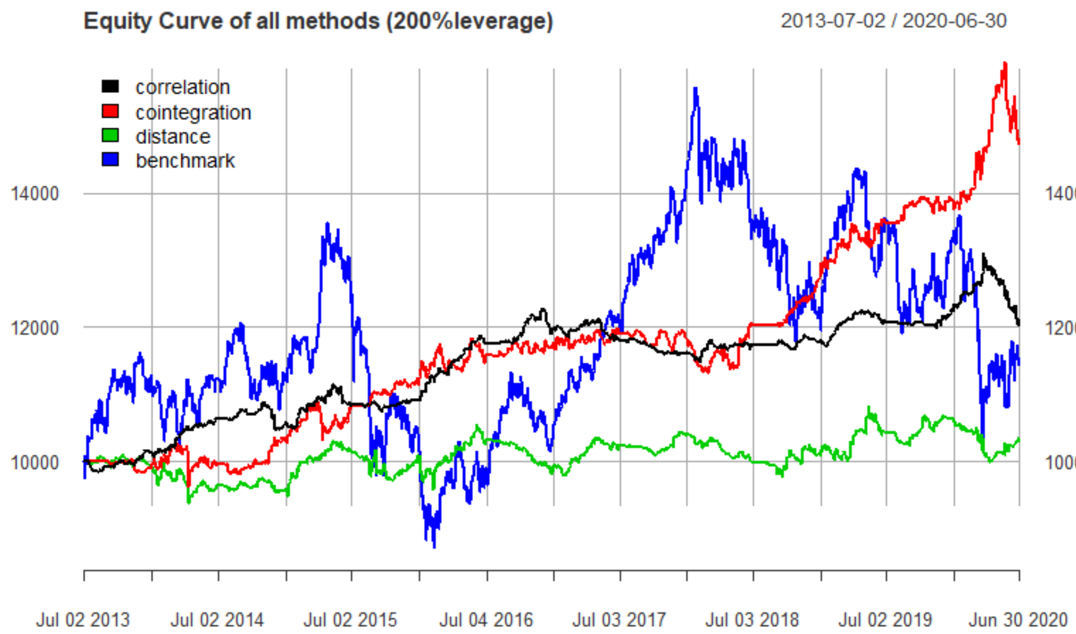
Table 6. Performance of all methods and benchmark for N, N/2 and N/3

Method	Financial Leverage	AR (%)	ASD (%)	ARC (%)	SP	MD (%)	IR
Correlation	HK$10000/N(100%)	10.13	**1.71**	1.42	-0.043	4.12	0.830
Cointegration	HK$10000/N (100%)	20.95	3.01	2.81	0.439	4.11	0.934
Distance	HK$10000/N (100%)	2.10	2.69	0.30	-0.442	**4.03**	0.113
Correlation	HK$10000/(N/2) (200%)	20.03	3.42	2.70	0.353	8.44	0.788
Cointegration	HK$10000/(N/2) (200%)	48.13	6.07	5.90	0.726	7.65	0.972
Distance	HK$10000/(N/2) (200%)	3.63	5.33	0.52	-0.182	7.78	0.098
Correlation	HK$10000/(N/3) (300%)	30.28	5.15	3.93	0.474	12.6	0.763
Cointegration	HK$10000/(N/3) (300%)	**82.55**	9.23	**9.17**	**0.833**	11.1	**0.994**
Distance	HK$10000/(N/3) (300%)	4.61	7.95	0.66	-0.104	11.5	0.083
Benchmark	-	15.05	17.89	2.07	0.032	35.7	0.115

Note: AR: Absolute return; ASD: Annualized Standard Deviation; ARC: Annualized Return; SP: Sharpe Ratio; MD: Maximum Drawdown; IR = ARC/ASD: Information Ratio. Bold font indicates the best results in case of each performance statistics.

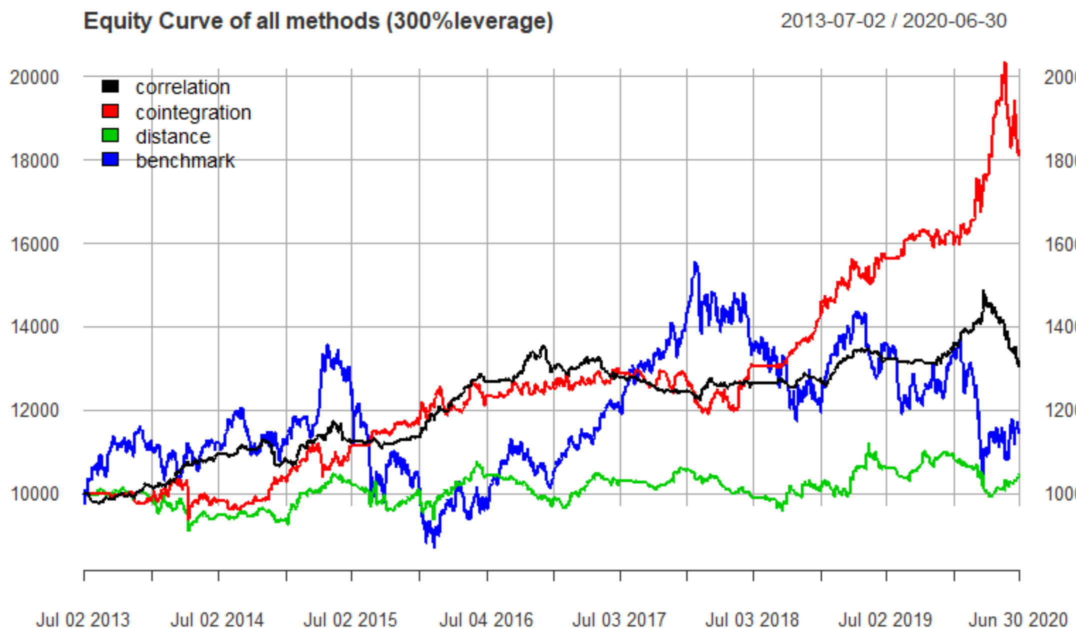
Figures 5 and 6 plots the equity curves of all three methods and the benchmark for the double degree of financial leverage and triple degree of financial leverage, respectively. Both graphs show that in the recession period, the correlation and the cointegration methods perform better than the benchmark while the distance method performs poorer. The distance method cannot outperform the other two methods in all market conditions. As we indicated above, their equity curves are characterized by much higher volatility, which is shown clearly in Figure 6.

Figure 5. Equity curve of all methods and benchmark – N/2



Note: The result of pair trading strategy with signals based on correlation, cointegration and distance methods compared with the results of Buy&Hold strategy for Hang Seng Index. The trading period started on 07/01/2013 and ended on 07/01/2020. The number of pairs traded was 10. The rebalancing period was 1 year. Degree of financial leverage = 200%.

Figure 6. Equity curve of all methods and benchmark – N/3



Note: The result of pair trading strategy with signals based on correlation, cointegration and distance methods compared with the results of Buy&Hold strategy for Hang Seng Index. The trading period started on 07/01/2013 and ended on 07/01/2020. The number of pairs traded was 10. The rebalancing period was 1 year. Degree of financial leverage = 300%.

Conclusion

This paper aims to test the profitability of the approach to pair selection for the purpose of pair trading strategy in the Hong Kong stock market. The main hypotheses of this paper are (RH1) *Is the pair trading strategy based on the distance, correlation, and cointegration method profitable in the Hong Kong stock market?* and (RH2) *Which method is a superior pair selection model for pair trading among three methods?* Based on these two hypotheses, the additional questions are constructed as follows: *Whether our results are robust to* (RQ1) *varying number of pairs selected;* (RQ2) *varying rebalancing period;* and (RQ3) *varying degree of financial leverage?*

The dataset used for this empirical research consisted of 50 Hang Seng index constituents listed in the document titled: *Hang Seng Indexes Announces Index Review Results* which was released usually in May from 2013 to 2020 by Hang Seng Indexes Company Limited. We also selected the Hang Seng index itself which was used as the benchmark for our strategy. The data is collected daily over the period from 07/01/2013 to 07/01/2020. The pairs are established from the pool of these 50 stocks. For the correlation method, we apply the Pearson correlation coefficient to estimate the correlation between 2 series of log returns, for the cointegration method, we employ Engle–Granger two-step approach to test for cointegration between 2 time series of prices, for the distance method, we first standardize the prices and then calculate the distance between 2 time series of prices. To generate buy and sell signals for pair trading strategy, we refer to Bui and Ślepaczuk (2020)'s approach, the volatility breakout model is applied with upper and lower thresholds calculated based on the exponential moving average and rolling standard deviation of the time series. The window size of the exponential moving average, the window size of rolling standard deviation, and multiplier m are parameters to be optimized. The performance of the distance method is compared with the benchmark (buy and hold Hang Seng index) as well as with the correlation and the cointegration methods in terms of absolute return, annualized return, annualized standard deviation, maximum drawdown, Sharpe Ratio, and Information Ratio.

Overall, the result shows that in the Hong Kong stock market, the pair trading strategy based on the distance, the correlation, and the cointegration method can obtain positive profits, "beating the benchmark" which indicates all three methods are effective. Among them, the cointegration method performs the best. The performance of the correlation method is superior as compared to the distance method. Moreover, pair trading strategy in case of all methods can outperform the market in recession stage, cannot in the expansion stage, but once again can in the longer term consisting of bull and bear markets periods, which confirms the market neutrality of pair trading strategy. However, our result is quite sensitive to the different number of pairs traded and rebalancing period and less sensitive to the financial leverage degree. For example, with 6 months rebalancing period, the distance method performs the best among the three methods evaluated. In specific, it is not clear whether the cointegration and the correlation method's performance improves with the greater number of pairs. On the other hand, all methods' returns and volatility are greater with a higher degree of financial leverage, hence the information ratio does not improve, except for the cointegration method.

To conclude, from the obtained result, we can state that the pair trading strategy based on the distance, correlation, and cointegration method is still profitable in the Hong Kong stock market (first research hypothesis) and the cointegration method of pair selection is superior as compared to the correlation method and the distance method (second research hypothesis). Furthermore, the obtained result is sensitive to varying number of pairs (first research question) and varying rebalancing periods (second research question) and less sensitive to varying degree of financial leverage (third research question).

References

Afawubo K. Pairs trading and selection methods: is cointegration superior?. Applied Economics, 2015, 47(6):599-613.

Bowen D, Hutchinson M C, O'Sullivan N. High Frequency Equity Pairs Trading: Transaction Costs, Speed of Execution and Patterns in Returns. Social Science Electronic Publishing, 2010, 5(3):31-38.

Bui Q., Ślepaczuk R., 2021, *Applying Hurst Exponent in Pair Trading Strategies* on Nasdaq 100 index, Physica A, https://doi.org/10.1016/j.physa.2021.126784 , https://www.sciencedirect.com/science/article/abs/pii/S037843712100964X,

Caldeira, J.F., Moura, G.V., Selection of a Portfolio of Pairs Based on Cointegration: A Statistical Arbitrage Strategy. Rev. Braz. Financ. 11(1), 49–80 (2013)

Chen, H., Chen, S.J., Li, F. (2012), Empirical Investigation of An Equity Pairs Trading Strategy. Working Paper, Columbia University.

Do, B. and Fa_ff, R. (2010). Does simple pairs trading still work? Financial Analysts Journal,66(4): 83-95.

Do, B. and Faff, R. (2012). Are pairs trading profits robust to trading costs? Journal of Financial Research, 35(2): 261-287.

Engle RF, Granger CWJ. Co-Integration and Error Correction: Representation, Estimation,and Testing[J]. Econometrica, 1987, 55(2): 251-276.

Fernando B S D S, Ziegelman F, Caldeira J. Pairs Trading: Optimizing via Mixed Copulaversus Distance Method for S&P 500 Assets. Social Science Electronic Publishing, 2017.

Gatev E, Goetzmann W N, Rouwenhorst K G. Pairs Trading: Performance of a Relative-Value Arbitrage Rule[J]. Social Science Electronic Publishing, 2006, 19(3): 797-827.

Hauke, J.; Kossowski, T. Comparison of values of Pearson's and Spearman's correlation coefficients on the same sets of data. Quaest. Geogr. 2011, 30, 87–93.

Hossein Rad & Rand Kwong Yew Low & Robert Faff, 2016. The profitability of pairs trading strategies: distance, cointegration and copula methods, Quantitative Finance, Taylor & Francis Journals, vol. 16(10), 1541-1558, October.

Huck N. The high sensitivity of pairs trading returns. Applied Economics Letters, 2013,20(14): 1301-1304.

Kijewski M., Ślepaczuk R., 2020, *Predicting prices of S&P500 index using classical methods and recurrent neural networks*, Working Papers of Faculty of Economic Sciences, University of Warsaw, WP 27/2020 (333), https://www.wne.uw.edu.pl/index.php/download_file/5769/4919/

Kość K., Sakowski P., Ślepaczuk R., 2019, *Momentum and Contrarian Effects on the Cryptocurrency Market*, Physica A 523, pp. 691-701, https://doi.org/10.1016/j.physa.2019.02.057, https://www.sciencedirect.com/science/article/pii/S037843711930216X?dgcid=author

J. Wang, C. Rostoker and A. Wagner, A high performance pair trading application., IEEE International Symposium on Parallel and Distributed Processing Symposium, 2009, 1-8.

Liew, R. Q., and Y. Wu. 2013. "Pairs Trading: A Copula Approach." Journal of Derivatives& Hedge Funds 19 (1): 12–30. doi:10.1057/jdhf.2013.1.

Lin Y X, Mccrae M, Gulati C. Loss protection in pairs trading through minimum profit bounds: a cointegration approach. Journal of Applied Mathematics &Decision Sciences, 2007, 2006(1).

Liu B, Chang L B, Geman H. Intraday pairs trading strategies on high frequency data: the case of oil companies. Quantitative Finance, 2016, 17(1): 1-14.

Puspaningrum H, Lin Y X, Gulati C M. Finding the optimal pre-set boundaries for pairstrading strategy based on cointegration technique. Journal of Statistical Theory & Practice,2010, 4(3): 391-419.

Ramos-Requena, J.P.,Trinidad-Segovia, J. E. and Sánchez-Granero, M. A. Introducing Hurst exponent in pair trading. Physica A: Statistical Mechanics and its Applications, Elsevier, vol. 488(C), 39-45.

Smith R T, Xu X. A good pair: alternative pairs-trading strategies. Financial Markets and Portfolio Management, 2017, 31(1): 1-26

Vidyamurthy G. PairsTrading: Quantitative Methods and Analysis / G. Vidyamurthy .Pearson Schweiz Ag, 2004: 35–47.

University of Warsaw
Faculty of Economic Sciences
44/50 Długa St.
00-241 Warsaw
www.wne.uw.edu.pl