American Journal of Business and Society
Vol. 5, No. 3, 2020, pp. 155-164
http://www.aiscience.org/journal/ajbs

A Profitable Stock Investment System Using the Cointegration Technique

Carol Anne Hargreaves*, Jariah Bte Abdul Nassar, Hu Yiming

Department of Statistics and Applied Probability, Faculty of Science, National University of Singapore, Singapore

Abstract

Pairs trading is a quantitative trading strategy that exploits financial markets that are out of equilibrium. The aim of this study is to identify stocks that behave in a similar manner, however, when there is a temporary divergence, there is an opportunity to make money. There are numerous ways to choose pairs of stocks to implement this trading strategy, the method we chose is cointegration. In the paper, we aim to show that the pairs trading strategy using cointegration is a robust method that consistently reaps profits across consecutive time periods as well as different sectors in the Australian stock market. Our results prove this; we saw great profits in our trading periods, with the greatest being a hefty 38%. The results further demonstrate that pairs trading is effective; profits were positive across two periods and three sectors. Furthermore, in 5 out of these 6 scenarios, our strategy made profits that are at least 2 times greater than the market profit.

Keywords

Stocks Analysis, Trading Strategies, Market Neutral Trading Strategy, Cointegration, Pairs Trading, Australian Stock Market

Received: August 21, 2020 / Accepted: September 9, 2020 / Published online: September 28, 2020

1. Introduction

The only study [1] that uses a pairs trading strategy and tested thoroughly using extensive datasets, a wide variety of securities and across different financial markets, was the pairs trading study based on the distance method, where closely correlated security prices were grouped as pairs and traded when their prices diverged by more than a pre-specified amount. In this study, we examined a pairs trading strategy based on the cointegration method and applied it to stocks in the Australian stock market. Pairs trading presented a positive performance in past studies [24,2] and this is one of the motivations for this study. We aimed to identify a pair of stocks in the same sector that is known to historically have a similar price movement and have a positive linear relationship over time. However, when there is a temporary divergence, the *outperforming* stock (the stock that moved up) would be sold and the *underperforming* stock (the stock that moved down) would be bought, with the understanding that the *spread* between the two stocks would eventually converge by either the *outperforming* stock moving back down or the *underperforming* stock moving back up or both, resulting in your trade making money in all of these scenarios. If both the stocks moved up or moved down together without changing the spread between them, you will not make or lose any money.

We identified pairwise stock correlations for all stocks in three chosen sectors. The top 3 pairs of stocks in each of these sectors form an investment stock portfolio with equal weightage. Once we had identified our portfolio for each sector, we implemented and executed a long-term trading strategy over a six-month time period. The returns from each of the six stock portfolios are compared to the Australian stock market index performance. We hypothesized that the six stock portfolios will outperform the Australian stock market Index. The cointegration-based pairs trading strategy was applied to two consecutive time periods to demonstrate the consistency and reliability of the strategy and to rule out

* Corresponding author
E-mail address: carol.hargreaves@nus.edu.sg (C. A. Hargreaves)

possible random results.

This paper is structured into 5 sections. While Section 1 is the introduction, Section 2 covers the data extraction process, Section 3, a brief overview of the methods used, Section 4 the analysis results, after which Section 5 presents the conclusion and ending with Section 6 where recommendations for future studies are discussed.

2. Data Extraction

We used the open-source R programming software (Version 3.6.0) for our study. We retrieved the Australian stock data for the chosen sectors from Yahoo! Finance [3] by using the "quantmod" package in R.

We extracted the adjusted close prices of all stocks in the chosen sector for our predetermined training period of three years. We needed the training period to be three years as cointegration is a long-term phenomenon and we needed sufficient historical data to be able to pick up on it. In particular, the 2 three-year training periods that were chosen were 18th December 2012 to 18th December 2015 and 31st May 2016 to 31st May 2019. We also extracted the 2 6-month testing periods to test and evaluate the performance of our cointegration method based on the profits derived from trading during the testing periods; 21st December 2015 to 25th May 2016 and 3rd June 2019 to 15th November 2019. Both our training and trading periods are long as stocks are relatively less risky in the long investment periods than in shorter periods [19].

Instead of considering all stocks for pairs trading, we selected stock pairs based on individual sectors. This is because, for pairs trading, we are interested in stocks that behave in a similar nature (moving in the same direction). Such pairs were identified through a series of statistical tests [23]. It is more likely for us to find stocks of similar nature within a sector rather than across the entire market [17]. Furthermore, sectors are an example of economically meaningful groups highlighted by [20] for pairs selection. Hence, at any one time, we are looking at a reduced stock universe [23] or sector. In addition to that, it was reported by [29] in their paper that trading within sector proved high levels of stability or low volatility.

Penny stocks have high liquidity costs [4] and are considered to highly volatile. To reduce our trading risk, we excluded stocks that were less than one dollar. In the next section, we introduce our pairs trading methodology.

3. Methodology

We deployed a simple methodology for our pairs trading strategy. A flow chart of the steps taken to identify a stock pair and to evaluate the trading performance of the stock pair is shown in figure 1 below.

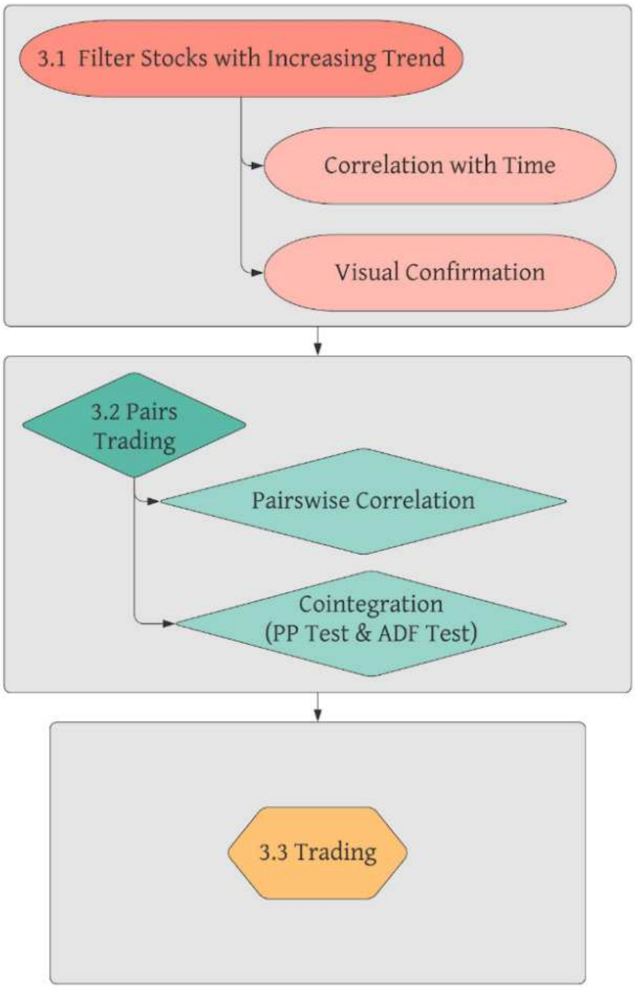

Figure 1. Flowchart for Methodology.

3.1. Selection of Stocks with an Upward Trend

The first step in our technical analysis approach was to identify stocks with an upward trend. To do this, we constructed a 2-layered filter, the first filter selected only those stocks whose prices had a strong positive significant correlation with respect to time over a 3-year period. The second filter selected stocks that appeared to have an upward trend using visual confirmation by manually confirming whether the stock price trends were upwards or not. For example, in figure 2 below, the stock MCY.AX had a high positive correlation coefficient of 0.819, however, we see that after 6th March 2015 (circled in red), there was a dip in the price. Hence, we do not consider the MCY.AX for further analysis, even though the correlation coefficient was as high as 0.819.

Figure 2. MCY.AX Prices in Period 1.

The correlation was obtained using the Pearson Correlation Coefficient [5] formula.

$$r = \frac{n(\sum xy) - (\sum x)(\sum y)}{\sqrt{[n\sum x^2 - (\sum x)^2][n\sum y^2 - (\sum y)^2]}} \quad (1)$$

The Pearson Correlation Coefficient (r) indicates the degree and direction of a linear association between two variables such that $-1 \leq r \leq 1$ [18]. A negative correlation coefficient indicates a negative linear relationship [18], this means that as one variable increases, the other variable decreases, and vice versa. A positive correlation coefficient denotes a positive linear relationship [18], this means that as one variable increases (decreases), the other variables increases (decreases) too. A correlation coefficient of zero indicates that the two variables do not exhibit a linear relationship [18], this means that a change in one variable is not accompanied by a change in the other variable. Hence, coefficients near zero have a weaker linear relationship between the variables whereas coefficients near -1 or 1 have a strong linear relationship.

In our case, we have only considered stocks whose prices had a positive correlation coefficient with respect to time of at least 0.8. Hence, in the above equation (1), we substituted the stock prices into x and values from 1 to n into y. n denotes the number of days in the training period. Note that y or "time" is increasing linearly from 1 to n. A stock that has increasing prices would exhibit a strong correlation with this "time" variable. An example can be seen in figure 3 below, in the first period, the stock APA.AX's price against time had a large positive correlation coefficient of 0.952. The continuously increasing prices are desirable. We disregarded negatively correlated stocks and stocks with weak correlation with respect to time as shown in figure 4. The stock CEN.AX's price against time has a low positive correlation coefficient of 0.0135. Hence, for further analysis, we considered APA.AX but not CEN.AX.

Figure 3. APA.AX Prices from 18th December 2012 to 18th December 2015.

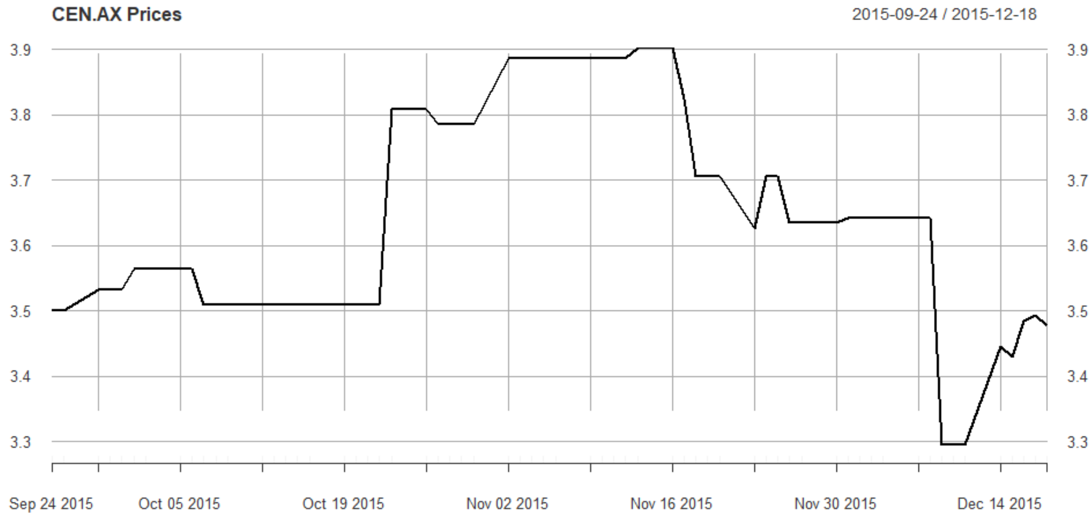

Figure 4. CEN.AX Prices from 18th December 2012 to 18th December 2015.

In addition to the Pearson correlation coefficient, we also computed the probability value (or p-value) of this statistic. The p-value is an indicator of whether we should or should not accept the null hypothesis presented by the statistic. The null hypothesis put forward by the Pearson correlation statistic is that there exists no significant linear relationship between the two variables [5]. Hence, when the corresponding p-value was smaller than 5%, we rejected the null hypothesis and concluded that the two variables had a linear relationship at the 5% significance level [5].

In the context of our study, stocks with a strictly increasing linear trend were selected for further analysis in the next section, if and only if:

1) The correlation coefficient of stocks' prices versus time was greater than 0.8 and

2) The corresponding p-value was smaller than 0.05

3.2. Pairs Trading

The origins of pairs trading can be found at [2, 6] and [25]. According to [6], if two stocks have similar characteristics (for example, Coca-Cola and Pepsi are both in the same market segment and are both likely to be affected by the same market events, such as the pricing of common ingredients, may be considered for pairs trading), then the prices of both stocks must be more or less the same and that if the prices happen to be different, it could be that one of the stocks is overpriced, or the other stock is underpriced, or the mispricing is a combination of both. Pairs trading involves selling the higher-priced stock and buying the lower-priced stock with the idea that the mispricing will correct itself in the future. The mutual mispricing between the two stocks is captured by the notion of spread. The greater the spread, the higher the magnitude of mispricing, and the greater the profit potential [6].

According to [7], pairs trading is a strategy that implements statistical arbitrage and convergence. The basic idea is mean reverting; prices tend to move back to the mean which holds advantageous implications for long term investors [19] like us. If two stocks can be identified that have a relatively high correlation, then the change in the difference in price between the two stocks can be used to signal trading opportunities if one or more of the two stocks move out of the correlation with the other stock [7, 21].

If the change in the spread between the two stocks exceeds a certain level (their correlation has decreased), then the higher-priced stock can be considered to be in a short position [21] and should be sold as it is assumed that the spread will decrease as the higher-priced stock returns to the mean (decrease in price as the correlation returns to a higher level). Likewise, the lower-priced stock is in a long position [21], and it is assumed that the price will rise as the correlation returns to normal levels. The buying and selling of equal amounts of the two stocks give us an advantage as we will profit as the two prices move back into correlation [7].

It is believed that the first academic trading paper was authored by [24]. They used distance to choose the stock pairs; a partner was chosen for each stock by taking the stock that minimizes the sum of squared deviations between the two normalized price series. The distance method was since then been critiqued to be non-parametric. Although it was observed that the stocks moved together in the past, there exists no guarantee that this phenomenon will continue to occur in the future [23].

When examining multivariate series to determine statistically if there is a cause-effect relationship between the variables represented by the time series, Engle and Granger observed that even though the two series are nonstationary, it is possible that in some instances a specific linear combination of the two is stationary, that is the two series move together. Engle and Granger coined the term co-integration [8].

However, works by authors [25] and [6] have failed to show any back-testing results along with their theoretical analysis. A study done by [26] attempts to generate results, however it deals with a small group of pairs that were selected in advance and not a broad market. Similarly, [2] carried out a high frequency pairs trading methodology in the highly liquid secondary market for U.S. government debt and not in a stock market.

3.2.1. Pairwise Correlation

We identified possible stock pairs by computing the pairwise Pearson correlation and p-value for all stocks in each sector. Pairs of stocks with correlation values greater than 0.8 and p-values less than 0.05 were considered stock pair candidates. This condition was motivated by [22] & [28]. For example, the stock pairs, (SCG.AX and GPT.AX) and (DOW.AX and ALQ.AX) exhibited strong pairwise correlations with correlation coefficients of 0.981 and 0.886, respectively From figures 5 and 6 below, we can also visually confirm that the prices of both stock pairs (SCG.AX and GPT.AX) and (DOW.AX and ALQ.AX) move in a similar manner. Hence, these stocks will be considered for pairs trading.

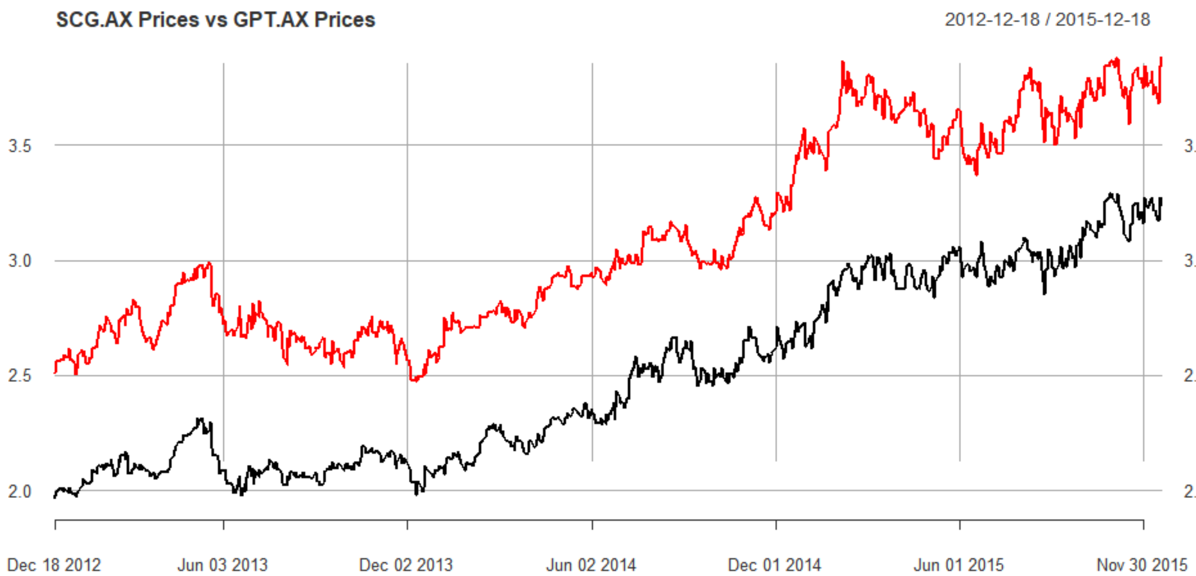

Figure 5. Co-movement of SCG.AX Prices with GPT.AX Prices in Period 1.

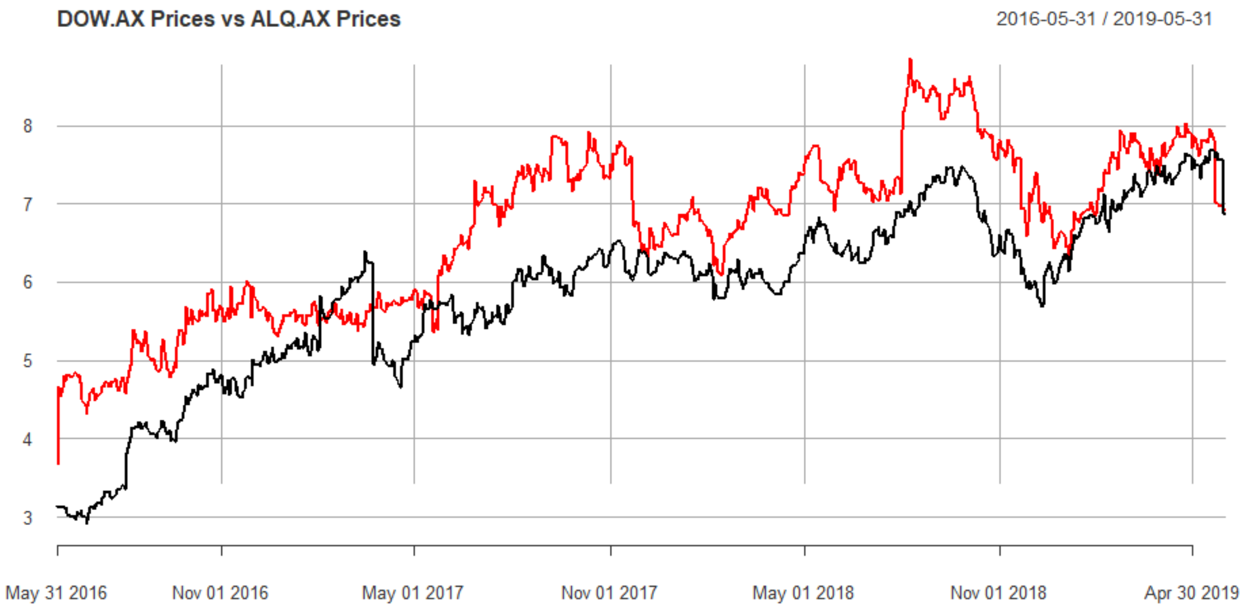

Figure 6. Co-movement of DOW.AX Prices with ALQ.AX Prices in Period 2.

3.2.2. Cointegration

There exist various methods to conduct pairs trading; correlation, distance, etcetera [9, 22]. In this paper, we implemented pairs trading using pairs of stocks that were cointegrated. In layman terms, cointegration can be thought of as a correlation in a longer period of time. In reality, this translates to the pair of stocks exhibiting similar behaviour (price movement) with a slight lag (one stock's price moves slower than the other). We take advantage of this lag. If one of the stocks, stock A is performing well at time X, we can be highly certain that after some time at time Y, the other stock, stock B will follow and be performing well too. In this case, we will sell shares of stock A [27] and buy shares of stock B at time X. At time Y, we can sell the shares of B that we hold and see a profit. The inverse is true; if A is performing poorly at time X, we can be sure that at time Y, B will be performing poorly as well. We will then take the necessary action.

Time series is defined to be data that is arranged in chronological order [10]. It is considered discrete data [10], where each value is numeric in nature. This definition is relevant to us as we considered the stocks' prices to be an example of a time series. There are many statistical derivations to be made from the stocks' prices when we consider them as time series [10]. From now, we will use stock prices and time series interchangeably.

Before moving on to Section 3.3 where we performed trading, we needed to identify pairs of stocks that are suitable for trading. To determine which pairs of stocks are appropriate, we performed a two-pronged approach [11] which helped us to conclude if pairs of stocks are cointegrated or not. This approach was inspired by the Engle-Granger Cointegration Model [11]. To conclude that the two time-series are cointegrated, we required the fulfilment of the following two conditions [11];

1. Individually, the time series are non-stationary or more specifically they are integrated of order 1; I(1). This can be understood as the time series being dependent on time.
2. A linear combination of two I(1) series is stationary or integrated of order 0. A stationary time series is one whose statistical properties such as mean, variance, etcetera is independent of time and are constant. This means that these statistical properties will not change over time which can be especially useful for making assumptions for the future.

To check that the above two conditions are satisfied for every pair of stocks, we made use of two stationarity tests, namely the Phillips-Perron Test and the Augmented Dickey-Fuller Test. The implementation of these tests is detailed below.

3.2.3. Phillips-Perron Test

To satisfy condition 1 stated above, we employed the use of the Phillips-Perron Test. This test postulates that the time series is of the form

$$\Delta y_t = (\alpha - 1)y_{t-1} + u_t \; (t = 1, 2, \dots) \tag{2}$$

The above equation posits that the stock's price at time t depends on the stock's price at time $t-1$ with a factor α [12]. Note that Δ is the first difference operator. This test puts forward the null hypothesis that a time series is integrated of order 1 or simply that $\alpha = 1$. In our analysis, we are more concerned with the statistical significance of this test rather than the test statistics themselves. Hence, we will look at the p-value associated with this test (similar to section 3.1). If the p-value is greater than 0.05, then we concluded that there is insufficient evidence to reject the null hypothesis. Hence, we accepted such a time series to be integrated of order 1. Hence, for our analysis, we kept these stocks and discarded stocks that return a Phillips-Perron test p-value smaller than 0.05.

3.2.4. Augmented Dickey-Fuller Test

For the fulfilment of condition 2 above, we needed to consider a linear combination of both time series. For this, we performed a linear regression of two stocks that have satisfied condition 1. Suppose we performed a linear regression of stock A's prices against B's, we will be dealing with the following equation [13];

$$A = \beta_0 + \beta_1 B \tag{3}$$

By fitting stock A's prices against stock B's, we find the slope β_0 and the intercept β_1. The most useful value here to us is β_1, we call this the hedge ratio. Moving forward, we constructed the spread

$$spread = A - \beta_1 B \tag{4}$$

In our analysis, we considered this spread as the linear combination of the two time series. To test if this spread (or the linear combination) was stationary, we employed the use of the Augmented Dickey-Fuller Test. This test proposes a similar model as seen in (2) but with an additional lag. It puts forward the null hypothesis that there exists a unit root in the time series; $\alpha = 1$ [14]. Since we required our time series to be stationary (integrated of order 0) and not contain any unit root, this condition will be met if the p-value is smaller than 0.05. Hence, for our analysis, we kept these pairs of stocks and discarded those that return an Augmented Dickey-Fuller test p-value greater than 0.05.

It should be noted that when dealing with two I(1) stocks, stock A and stock B, we perform a linear regression of A against B and B against A, and test both spreads for stationarity. If both spreads are stationary, we chose the spread with the smaller hedge ratio. If only one of the two spreads was stationary, we chose that spread. Otherwise, if both spreads were not stationary, we moved on to our next step without taking into consideration this pair of stocks.

3.3. Trading

Pairs of stocks that are identified to be cointegrated based on the three-year historical data are traded for the next six months. We traded with an initial investment of $10000 and assumed a flat transaction cost of $25.

To help us decide when to trade, we created trading signals. We used an upper and a lower signal that corresponded to 1.25 standard deviations away from the mean of the spread calculated according to the training data. This was motivated by [24] and [15].

When we're in the trading period, the prices of the pair of stocks; stock A and B are updated daily except on weekends. Next, we computed the daily spread, by using the hedge ratio of stocks A and B that were calculated based on the training data in [3].

When the new spread was above the upper signal, it meant that stock A was performing better than the hedged stock B. Therefore, we bought shares of hedged stock B and sold shares of stock A if we were holding any. Similarly, when the new spread was below the lower signal, it translated to hedged stock B performing better than stock A. Hence, we bought shares of stock A and sold shares of hedged stock B if we were holding any.

On the last day of the trading period, we sold all the shares that we had held as proposed by [24].

4. Results and Findings

We completed the full analysis for three sectors in two periods. To showcase the results of our methodology in detail, we will only illustrate the methodology and results of Healthcare sector in period 2.

4.1. Pairwise Correlation

The healthcare sector contained 27 non-penny stocks, 8 of which had strictly increasing linear prices. Out of the 28 possible pairwise combinations of stocks, 23 of them were found to have large and significant positive pairwise correlation coefficients.

4.2. Testing for Cointegration

We applied the Phillips-Perron Test [12] and the Augmented Dickey-Fuller Test [14] as specified in section 3.2.2. Two pairs of stocks were found to be cointegrated in the Healthcare sector. They were RMD.AX and CSL.AX and, FPH.AX and SHL.AX. All four stocks were confirmed to be non-stationary by the results of the Phillips-Perron Test as seen in table 1 below. The Phillips-Perron Test p-values were greater than 0.05 for all four stocks. Hence, there was insufficient evidence to reject the null hypothesis that RMD.AX, CSL.AX, FPH.AX or SHL.AX is I(1). Hence, we will keep these stocks for the Augmented Dickey-Fuller Test.

Table 1. Phillips-Perron Test p–values.

Stocks	Phillips Perron Test P-Value
RMD.AX	0.0996683
CSL.AX	0.3430864
FPH.AX	0.1822168
SHL.AX	0.1985666

When we performed the Augmented Dickey-Fuller Tests on the spreads we considered both spreads to determine which pair of stocks was the best choice. The results of the four spreads are shown in table 2 below.

Table 2. Augmented Dickey-Fuller p-values and Hedge Ratios for Pairs of Stocks.

Spread	Augmented Dickey-Fuller Test P-Value	Hedge Ratio, β
RMD.AX vs CSL.AX $RMD - \beta\ CSL$	0.01730144	0.07163334
CSL.AX vs RMD.AX $CSL - \beta\ RMD$	0.01938313	13.14395
FPH.AX vs SHL.AX $FPH - \beta\ SHL$	0.03924011	0.9798622
SHL.AX vs FPH.AX $SHL - \beta\ FPH$	0.0528252	0.8010549

Looking at the RMD.AX and CSL.AX pair, we found that the augmented Dickey-Fuller tests returned p-values less than 0.05 for both spreads, indicating that they were both stationary. Therefore, we used the hedge ratio to choose our spread. Since, RMD.AX against CSL.AX yielded a smaller hedge ratio, we chose it. We note that in this case, CSL.AX is the hedged stock.

For the pair FPH.AX and SHL.AX, one of the spreads; SHL.AX against FPH.AX returned a p-value greater than 0.05, indicating there is insufficient evidence to conclude that the spread was stationary. Hence, in this case, we chose the FPH.AX against SHL.AX spread where SHL.AX is the hedged stock.

4.3. Creating Portfolio (Trading)

After obtaining our list of cointegrated pairs of stocks, we arranged our stock pairs in a specific order that allowed us to select the top 3 pairs of stocks for our portfolio.

In this paper, we propose that we arrange the stocks in decreasing order of their pairwise correlation based on the training data. We already have computed these pairwise correlation coefficients and the corresponding p-values in section 4.1.1. Hence, it was natural for us to use this as a means of sorting our pairs of stocks. We select our top 3 pairs of stocks from this ordered list to form our portfolio.

We note that each pair of stock has equal weightage in the portfolio. We further note that if there are less than or equal to 3 pairs of cointegrated stocks, we move forward to trading without omitting any of them.

In our illustration of the Healthcare sector, we observe from section 4.2. that we only have two cointegrated pairs of stocks; RMD.AX against CSL.AX and FPH.AX against SHL.AX. Hence, these two pairs of stocks were placed in our portfolio.

4.4. Creating Signals

As discussed in section 3.3., we required trading signals to let us know when to buy, sell, or hold the stock. Based on our historical data we created the upper and lower signals. After creating the trading signals, we proceeded to trade for 6 months. For illustration purposes, we will only look at the trading of the RMD.AX and CSL.AX pair.

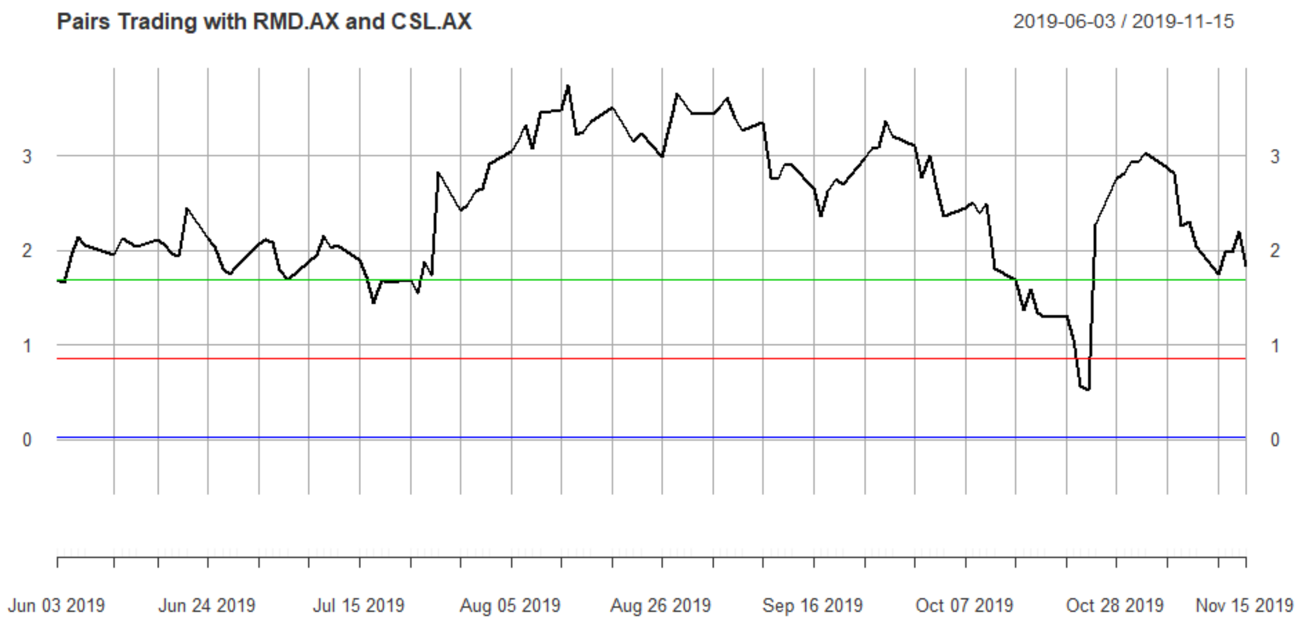

Figure 7. Trading Signals for RMD.AX against CSL.AX.

In figure 7, the black line is the spread of RMD.AX against CSL.AX in the trading period and the hedge ratio is 0.07163334 (computed in section 4.2.).

The red line is the mean of the spread in the training period. This was calculated to be 0.8599362. The upper and lower signals are the green and blue lines respectively. They were calculated to be 1.684722 and 0.03514996 respectively.

The trading was done by observing where the spread (black line) lies with respect to the signals. If it lies above the green line (upper signal), RMD.AX is performing better than CSL.AX and we should sell any shares of RMD.AX that we are holding and buy shares of the hedged CSL.AX. On the other hand, if the spread lies below the blue line (lower signal), it indicates that the hedged CSL.AX is performing better and we should sell any shares of the hedged CSL.AX and buy shares of RMD.AX. However, when the spread lies in between the upper (green) and lower (blue) signals, it means that the spread has not deviated much from the mean (neither stock is outperforming the other) and thus there is no point in trading during this time.

At the start, our spread (black line) lies above the upper signal (green line); RMD.AX performed better than the hedged CSL.AX, hence we bought shares of the hedged CSL.AX. In this particular example, interestingly it is observed that the spread never diverges beyond the lower signal (blue line). Hence, there is no change in action.

Table 3 below summarizes the pair trading transaction process in more detail. Days, where no transaction takes place, are omitted for conciseness.

Table 3. Trading of RMD.AX and CSL.AX Pair.

	Is Spread Smaller than Lower Bound?	Is Spread Greater than Upper Bound?	Action	Transaction	
03/06/2019	False	True	Buy Hedged CSL.AX	686	Shares of Hedged CSL.AX
15/11/2019	False	True	Sell Hedged CIM.AX & Buy WLL.AX	13450	Cash Value

From table 3, we can see that the final cash value was $13,450 which amounted to a profit of $3450 or 34%. Note that 15th November 2019 was the last day of trading, hence despite whatever action was recommended by the algorithm, we sold all stocks that we were holding.

In this trading period, the stock market index had a profit of 7%. This was calculated using the simple trading method; buying shares of the stock at the beginning of the test period and selling them at the end. Hence, by comparison, we can see that our method of pairs trading is working very well; the profit was almost five times higher.

4.5. Trading Summary

In table 4 below, displays the profits of the portfolios in the three sectors across the two time periods. A total of six stock portfolios.

Table 4. Comparison of Pairs Trading Portfolio against Market Index.

Sector & Period	Profit	Market Index
Finance 1	16%	
Healthcare 1	28%	5%
Industrial 1	38%	
Healthcare 2	35%	
Industrial 2	6%	7%
Utilities 2	14%	

From table 4, out of the six stock portfolios, five of the stock portfolios outperformed the stock market index and have a much greater return on investment than the stock market index return on investment, at least double with the greatest being eight times greater. There is only one sector; the Industrial sector in the second period that yielded fewer returns than the stock market index. Even so, it was only 1% less; a profit of $600 (6% of initial investment) was still realized.

5. Conclusion

Most historical studies on pairs trading were performed using a long-term study in the US market. In this paper, we focused on pairs trading using a short-term study in the Australian market. Although [16] implemented a cointegration pairs trading strategy, their empirical analysis only examined two Australian stocks over a sample period of one year. The main objective of this study was to find pairs of stocks that will outperform the stock market index in three different sectors over two non-overlapping time periods in the Australian Stock Market. We wanted to determine whether the identification of a cointegrated pair will increase in profit over a six-month period. The results in Section 4 demonstrated outstanding potential profitability for our cointegration-based pairs trading strategy. Out of 6 time periods, all six time periods delivered a return of at least 6%. The average profit across the two six-month periods was 23%.

In conclusion, pairs trading strategies continue to be a profitable trading strategy that remains robust (we used two non-overlapping trading periods and 3 different sectors) in highly volatile market conditions (our portfolios consistently outperformed the market index). The overall performance of our pairs trading strategy for the two six-month periods was almost four times more than the Australian Market Index. The overall return on investment was 23% and the Australian Market performance was only 6%.

6. Further Improvements

Further studies are recommended to determine whether our cointegrated pairs trading strategy can be extrapolated to other stock markets worldwide. We applied the cointegrated pairs trading strategy to a six-month trading period, it is suggested that other time periods such as 3 months, etc. may be considered for further study. In this study, we extensively used the mean-reverting phenomenon to our advantage. With further research, more discoveries on the length of the mean-reversion period could be made which could be particularly useful in creating portfolios for short-term and long-term investments.

In this study, each pair of stocks in the portfolio was given equal weight, further studies may explore investing in stocks by allocating different weights based on the expected return and risk of each stock. We believe that the combination of portfolio theory with our pairs trading method has great potential for increased profitability. The selection of stocks could also be made more stringent with additional conditions with the employment of index models.

References

[1] Andrade, S., Di Pietro, V. and Seasholes, M. “Understanding the profitability of pairs trading. Unpublished working paper, UC Berkeley, Northern University, 2005.

[2] Nath, P. (2003) High Frequency Pairs Trading with US Treasury Securities: Risks and Rewards for Hedge Funds. Working Paper, London Business School.

[3] Yahoo Finance – stock market live quotes, business & finance news. (n.d.). Retrieved December 11, 2019, from https://au.finance.yahoo.com/

[4] Liu, Q., Rhee, S. G., & Zhang, L. (2011). On the Trading Profitability of Penny Stocks. *SSRN Electronic Journal*. doi: 10.2139/ssrn.1917300

[5] E. I., O., Chikweru Amadi, E. (2018). Test for Significance of Pearson's Correlation Coefficient ( ). *ResearchGate*. Retrieved from https://www.researchgate.net/publication/323522779_Test_for_Significance_of_Pearson's_Correlation_Coefficient

[6] Vidyamurthy, G. (2004) Pairs Trading: Quantitative Methods and Analysis. New York: John Wiley & Sons.

[7] Heydt, Michael (2015). Mastering pandas for Finance. Packt Publishing.

[8] Engle, Robert F. and Granger, C. W. "Co-integration and Error Correction: Representation, Estimation and Testing." Econometrica 55, No. 2 (March 1987): 251-276.

[9] Blázquez, M. C., Carmen De La Orden De La Cruz, & Román, C. P. (2018). Pairs trading techniques: An empirical contrast. *European Research on Management and Business Economics*, *24* (3), 160–167. doi: 10.1016/j.iedeen.2018.05.002

[10] Kirchgssner, G., & Wolters, J. (2014). *Introduction to modern time series analysis*. Springer.

[11] Maddala, G. S., & Kim, I.-M. (2007). *Unit roots, cointegration and structural change*. Cambridge: Cambridge Univ. Press.

[12] Phillips, P. C. B., & Perron, P. (1988). Testing for a unit root in time series regression. *Biometrika*, *75* (2), 335–346. doi: 10.1093/biomet/75.2.335

[13] Montgomery, D. C., Peck, E. A., Vining, G. G., Montgomery, D. C., & Montgomery, D. C. (2012). *Introduction to linear regression analysis*. Hoboken, N. J: Wiley.

[14] Said, S. E., & Dickey, D. A. (1984). Testing for unit roots in autoregressive-moving average models of unknown order. *Biometrika*, *71* (3), 599–607. doi: 10.1093/biomet/71.3.599

[15] Jurek, J. W., & Yang, H. (2007). Dynamic Portfolio Selection in Arbitrage. *SSRN Electronic Journal*. doi: 10.2139/ssrn.882536

[16] Lin, Y., McCrae, M. and Gulati, C., "Loss protection in pairs trading through minimum profit bounds: A cointegration approach. Adv. Decis. Sci., 2006

[17] Elton, E., Gruber, M., Brown, S., & Goetzmann, W. (2014). *Modern Portfolio Theory and Investment Analysis, 9th Edition*. John Wiley & Sons.

[18] Egghe, L., Leydesdorff, L. (2009). The relation between Pearson's correlation coefficient r and Salton's cosine measure. *Journal of the American Society for Information Science and Technology.* Retrieved from https://arxiv.org/ftp/arxiv/papers/0911/0911.1318.pdf

[19] Spierdijk, L., & Bikker, J. A. (2012). Mean Reversion in Stock Prices: Implications for Long-Term Investors. *SSRN Electronic Journal*. doi: 10.2139/ssrn.2046093

[20] Mashele, H., Terblanche, S., & Venter, J. (2013). Pairs trading on the Johannesburg Stock Exchange. *Investment Analysts Journal*, *42* (78), 13–26. doi: 10.1080/10293523.2013.11082559

[21] Huang, C.-F., Hsu, C.-J., Chen, C.-C., Chang, B. R., & Li, C.-A. (2015). An Intelligent Model for Pairs Trading Using Genetic Algorithms. *Computational Intelligence and Neuroscience*, *2015*, 1–10. doi: 10.1155/2015/939606

[22] Tsoku, J. T., & Moroke, N. D. (2018). Pairs trading in JSE financial sector. *Journal of Statistics and Management Systems*, *21* (5), 877–899. doi: 10.1080/09720510.2018.1467647

[23] Harlacher, Markus (2016). Cointegration Based Algorithmic Pairs Trading. *(Unpublished doctorate thesis.)* University of St. Gallen, School of Management, Economics, Law, Social Sciences and International Affairs

[24] Gatev, E., W. Goetzmann, and K. Rouwenhorst (2006). Pairs trading: Performance of a relative-value arbitrage rule. Review of Financial Studies 19 (3), 797–827.

[25] Wilmott, P. (2006). Paul Wilmott on quantitative finance. Wiley Chichester.

[26] Do, B., R. Faff, and K. Hamza (2006). A new approach to modeling and estimation for pairs trading. In Proceedings of 2006 Financial Management Association European Conference.

[27] Jacobs, B. I. and Levy, K. N. (1993). Long/short equity investing. The Journal of Portfolio Management, Vol. 20, 1, pp. 52-63.

[28] Perlin, M. S. (2009). Evaluation of Pairs Trading Strategy at the Brazilian Financial Market. Journal of Derivatives & Hedge Funds, Vol. 15, pp. 122-136.

[29] Caldeira, J. and Moura, G. (2013). Selection of a Portfolio of Pairs Based on Cointegration: A Statistical Arbitrage Strategy. Available at SSRN 2196391.