Speed of Adjustment Filtering in Pairs Trading

George Bresnahan

Supervisor: Ian Corrick

Student No. 200406026

Word Count – 10998

This study explores the optimisation of Pairs trading through filtering candidate pairs based on the speed of adjustment coefficient obtained from the Johansen method of Cointegration testing (1988). With aim to assess how this modification affects the profitability of the trading strategy and to investigate the suggested cause for declining profits attributed to pair non-convergence in the existing literature. Empirical testing of this modification has been conducted using data from the Do and Faff assessed period (2003-2009) and recent years (2017-2024) using S&P 500 data. The results provide conclusions on the impact of speed of adjustment filtering and provide insights into the key determinants of the strategies profitability.

Introduction

Pioneered by Nunzio Tartaglia and his team of quantitative analysts at Morgan Stanley in the mid-1980s, ‘Pairs Trading’ a statistical arbitrage trading strategy was theorized and established. The strategy makes use of the distance between prices of two stocks that are regarded to have historically moved together, or comoved, in price. The strategy profits on short-run deviations (disequilibrium) from a long-run constant (equilibrium) asset price relationship. When these stocks diverge, an arbitrageur will short the stock with the higher share price and buy the stock with the lower share price. As these two stocks hold a long-run equilibrium when they converge, in theory, the arbitrageur will make profit from both the short and long positions. Since the strategy is dependent upon the price distance, spread, between the two asset prices rather than broad market movement, the strategy is hedged to market risk, and so is market neutral. This, among other reasons, makes Pairs Trading commonly used as a strategy for institutional investors.

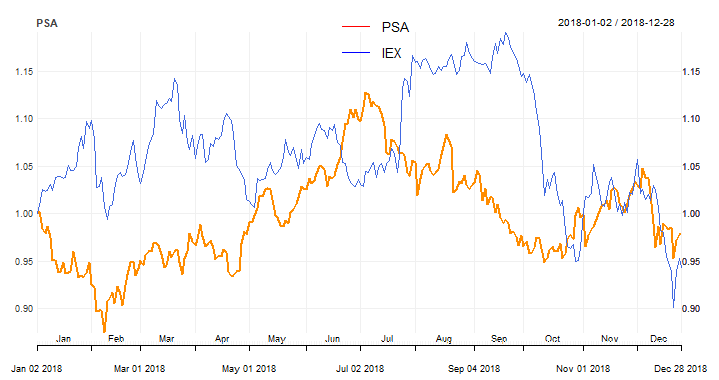
Figure 1

Figure 1 illustrates the trading strategy in practice. Using the example of Public Storage (PSA) and IDEX Corporation (IEX), a traded pair which yielded 12.7% returns in the year 2018, the long-term equilibrium relationship between the stocks is evident from the intertwining of their stock price time series. Trading opportunities become apparent when the stocks diverge, providing a practical illustration of the trading strategy.

In Pairs Trading, the process of pair matching to identify suitable pairs for trading is fundamental to the trading strategy. This aspect of the strategy is a large area of discussion in the literature, as later explored in the literature review. Two methods of pair creation are prominent; the Distance Method and the Cointegration Method, the former is the original approach, while the latter is an application of Econometrics to the pair creation process, introduced years after the trading strategy was initially characterized.

Despite such innovation of the pair creation process, academic literature has revealed a decline in profits from Pairs Trading since introduction. Moreover, academic contributions have signified the strategies’ profits lack robustness to trading costs. In response to these unfavourable outcomes, this dissertation introduces a new adaptation of the Cointegration Method of Pair Creation. The adaptation filters candidate pairs for trading based on the Speed of Adjustment Coefficient obtained from the Johansen method of Cointegration testing. This dissertation empirically investigates whether the new adaptation of the pair creation process significantly alters the profitability assessment of the widely used trading strategy.

Literature Review

The seminal piece of research on the strategy was conducted by Gatev, Goetzmann and Rouwenhorst (1999). GGR’s methodology consisted of taking daily data via CRSP, over the period of 1962-1997, from the U.S equity market and applying a pairs trading strategy to data. The trading strategy consisted of two stages, the formation period of pairs over a twelve-month timespan and the trading period, lasting 6 months. These stage lengths go on to be varied and assessed for profitability in later academic studies by Huck (2013) and Huck and Afawubo (2014). Pairs were selected by the matching process of minimizing the sum of the square deviations between two stocks normalized price series, suggesting that traders tend to choose pairs of stocks for trading that have historically exhibited close price movements**.** This method of matching became known in the Pairs trading literature as the ‘Distance Method' of pair creation. Trading is done according to a pre-determined rule, a long-short position is opened when the pairs' prices diverge by a certain amount, in GGR the open threshold selected was 2 historical standard deviations of the spread. The position is then closed when the pair converge and cross over. All positions are closed at the end of the 6-month trading period and pairs that have not crossed are accounted. Since pairs can diverge and converge multiple times during the trading period, it is important to consider the method for calculating excess returns. The returns are considered excess on the return from investing a dollar into each individual position. GGR calculated excess returns through the return on actual employed capital, which is the sum of the pair payoffs divided by the number of pairs that open during the trading period.

For analysis, the pairs were categorized into three subsets: the top 5, top 20 & top 20 after 100 pairs, the ranking based on the size of the historical distance measure, number one being the pair with the smallest historical distance. A profitability comparison was conducted, varying whether positions were opened or closed on the day of divergence/convergence, or the day after. The profits expectedly were lower when waiting for a day to alter holding positions rather than an instant action on the same day, illustrating the importance of speed in the trading strategy.

GGR found annual percentage excess returns on employed capital on the three subsets, by doubling the 6-month trading period, these returns were 12.02, 12.12 & 9.12 whilst mean excess returns for all pairs traded was 8.46%. Indicating the profitability of the strategy and the importance of finding optimal pairs to maximize returns. When computed with trading costs these profits fell significantly to a range of 3.36% to 1.76%. Despite this, the higher gross returns compared to the S&P 500, with potential for optimization, motivated further research.

Following this work, academic literature branched out, with particular focus around methods of Pair creation, another foundational method was theorized by Vidyamurthy (2004). However, for coherence this literature review will remain focused on the strand of literature based upon the Distance Method, DM, proposed by GGR (1999).

The next noteworthy piece of literature in the DM strain is the second edition of the seminal paper by Gatev, Goetzmann and Rouwenhorst (2006). GGR modified their analysis to assess if the profitability of pairs trading was changing with time, the results showed average monthly excess returns of 1.14 % with the top 5 & top 20 pairs achieving 1.3 & 1.4%. The new analysis of sub-period profitability showed that mean returns in the data set of 1963-1988 were four times that of corresponding returns in the period 1988 – 2002, this prompted research into why there were signs of decreasing returns and whether this trend was going to persist into the future.

In response to these questions, Do and Faff (2010) published the next important research in the strand of literature. The methodology of GGR was repeated but with addition of CRSP US equity market data from January 2003 – January 2009. Results were computed and sub-period analysis was taken on the periods 1962-1988, 1989-2002 & 2003-2009, with the aim of assessing excess returns on the Top 20 pairs by subperiod, and the number of non-convergent pairs in each subperiod. Under the delayed trading rule, of waiting a day to alter holding positions, the results showed a notable decline in average monthly excess return, profits went from 0.86% in period 1 to 0.37% in period 2, representing a 57% decrease. Subsequently, average monthly returns further decreased by 35% in period 3, over the 6 years following GGR’s sample, reaching 0.24% per month. During the later timeframe, the strategy frequently yielded negative monthly returns, starting in the 90s and continuing into the 21st century. The authors made sense of this through cross-sectional analysis of the behavior of pairs that were traded. Throughout the sample for each subperiod the number of nonconvergent pairs increased from 26% in period 1 to 39 & 40% in the following periods, showing that increasing nonconvergent pairs reduces profitability, as expected given returns are derived from convergence. To solve this, Do and Faff proposed the use of an alternative pairs matching algorithm to enhance profitability by minimizing nonconvergent trades, believing that the Distance Method did not suffice in efficiently matching pairs due to its high percentage of nonconverging pairs.

In 1974 publication of the spurious regression problem (Granger and Newbold, 1974) shed light on widespread econometric error. It was identified that linearly combined nonstationary variables were showing misleading statistical evidence of a linear relationship between the two variables, this was a large problem for economic interpretation. Engle and Granger (1987) solved this through introduction of a groundbreaking concept for which Granger would win a Nobel prize. It was found that non-stationary variables can be linearly combined to form a stationary process if they share an underlying stochastic trend. The variables are defined to be cointegrated with each other meaning, by definition, they have a long-run equilibrium relationship.

In the years following the concept's introduction, application of Cointegration-based trading was tested for suitability on various commodity spreads. Notably among the research; the crack spread (Girma and Paulson, 1999), the soy crush spread (Simon, 1999) and the spark spread (Emery and Liu, 2002). Employing Engle-Granger's two-step methodology, each variable within these distinct spreads were found to be cointegrated. This characteristic, which showed reversion of the variables in the spread to a long-run equilibrium allowed the authors to arbitrage on these commodities. Simulations in the papers of simple trading strategies were conducted yielding significant profits. Cointegration-based trading’s practicality and profitability were evident, the characteristic of a long-run equilibrium relationship making it useful for mean reversion trading strategies.

After the publication of pairs trading by GGR in 1999 it was apparent that Cointegration could be applied to the Wall Street trading strategy, first to formalize this was Vidyamurthy (2004) who published a book on the application to pairs trading describing the methodology to create pairs and the foundation of a Cointegration based trading strategy. The strategy was split into three steps, identifying potential stock pairs, Cointegration testing and trading rule design. Vidyamurthy emphasized the point that a sample of 5000 stocks, would generate ~ 12 million pairs, running Cointegration tests on the entire dataset is not feasible. His solution was to use; the Sum of Squares of Deviation, SSD, between stocks, to eliminate pairs that were unlikely to be cointegrated. If a pair of stocks are consistently far apart in stock price over time it is reasonable to assume they aren’t cointegrated. This identification of potential pairs to test was a point of interest in later studies using the Cointegration method, with various methods of screening used. The largest empirical Cointegration pairs trading research (Rad, Low and Faff, 2015) followed Vidyamurthy’s method. Other screening methods in the literature include comparing similarity in stock returns (Huck and Afawubo, 2014), using correlation between pairs as a screening method (Miao, 2014), or simply reducing the number of stocks tested through sector-specific research. These approaches follow the idea that it is simply not feasible to exhaustively test a large data set for cointegrated pairs.

The research ensuing Vidyamurthy (2004) focused on answering whether the Cointegration method of pair creation solved the declining profitability of pair trading found by GGR (2006) and Do and Faff (2010). The aim was to see if the problem of pair non-convergence could be solved by creating pairs that are cointegrated. The first work to assess this was by Caldeira and Moura (2013) which assessed profitability of the method in the Sao Paulo stock exchange during 2005-2012. Whilst Gulati, Lin, and McCrae (2006) assessed the profit per trade of the Cointegration approach, analysis of only two pairs of Australian stock over a narrow sample of 2 years means that this research cannot be deemed an indicative analysis of the method’s profitability. Gulati, Lin, and McCrae did however introduce the important idea of trading rules, in the form of minimum profit per trade, which is the earliest method of optimizing the pairs trading strategy. Caldeira and Moura collected 20 pairs with the largest Sharpe Ratio over the sample period, of the first year, to be traded out of sample. The results showed an exceptional annual mean return of 16.38% (15.87 post transaction costs), this was the first empirical evidence of the profitability of the Cointegration method. However, trading only took place over 7 years, and larger-scale research would have to be done to truly assess the Cointegration methods' profitability, with comparison to the DM.

The first comparison between the two methods of pair creation was conducted by Huck & Afawubo (2015), data was taken from the S&P 500 over the years 2000-2011, two parametrizations were used to see if there was an impact on profit; the number of standard deviations were considered as the opening trigger (2 vs 3), and the length of the formation period (12 months following GGR vs 24 months). Excess returns were calculated, and the pairs characteristics decomposed. Across all variations of the two parameters, Cointegration yielded higher excess returns by multiples of the DM. Specifically, Cointegration was most profitable with a 1-year formation period and an opening trigger of 3 standard deviations. For a 2 standard deviation opening the proportion of non-convergent pairs (non-intersecting pairs) were higher for Cointegration by 15% compared to the DM. Meaning that the DM had proportionally fewer non-convergent pairs, contrary to theory, does this mean that Cointegration doesn’t solve the profitability problem of non-converging pairs identified by Do and Faff? Whilst there were more non-convergent pairs than the DM, the percentage of these non-convergent pairs that were still profitable was 20% points higher for Cointegration. Showing, that whilst pairs may not converge, they still partially converge and are still profitable. In this study, Cointegration is a lower risk, of zero return, method of pair creation.

All results were symmetrical in the 2-year formation length adaptation of the methodology; however, profits were lower for Cointegration and, oppositely, were higher for the DM, implying that GGR and Do & Faff could have increased returns using a longer formation period. This implies that Cointegration pair creation is less profitable when applied using a longer period of formation. Reasons for this aren’t presented, however observing the results shows increasing non-traded pairs and significantly fewer profitable non-convergent pairs created from a 2-year formation period.

In the work of GGR there was debate as to whether the opening trigger should be higher than two standard deviations, was GGR cutting their profits by opening too early? The results showed for DM, following the GGR methodology, monthly excess returns were lower when using a trigger of 3 SD, alongside this the number of non-convergent pairs increased significantly; explaining the decrease in returns and supporting intuitions of Do Faff and GGR.

In the case of Cointegration, remarkably, returns nearly tripled. Accompanying this, the number of non-convergent pairs decreased by a large 15%. This logically supports the theory that the further apart comoving pairs are when a position is opened, the less likely they are to diverge further or remain as far apart. Thus, Cointegration can be assessed as effective for creating pairs and trading them using a wider opening trigger. The mathematical equilibrium relationship that characterizes cointegrated pairs must allow for positions to be opened with a higher assurance of achieving profitable convergence, especially when pairs are relatively far apart, in comparison to the DM. As a result of this if pairs converge over a greater distance, then, by construction of the trading method, more profit is made. This explains the much greater returns for Cointegration than the DM. In summary, Cointegration appears to be a more reliable method of creating pairs than the DM, which allows for higher pair volatility and consequently higher excess returns. This was supported by work of Blazquez et al (2018) who plotted the residual series, distance between a stock pair over time, of select American banks, and found through residual series analysis that ‘Cointegration chooses stocks in a more accurate and complete way than distance and is therefore preferred’.

The larger scale search that was needed to truly assess Cointegrations profitability was conducted by Rad, Low and Faff (2015) on the entire US equity market from 1962-2014, employing both the Distance and Cointegration methods. Contrary to Huck & Afawubo’s findings, these results show the DM outperforming Cointegration. Although Cointegration identified ~10% more pairs, lower excess returns are yielded from these pairs, 4.56% (DM), 3.96% (CI), and a higher proportion of negative returns (31% vs 29% for DM). Whereas Huck and Afawubo had recorded returns of Cointegration in the sample triple that of the DM.

Making sense of this, the methodology of each paper must be compared. Doing this a clear distinction can be identified in the potential pair identification procedure; for Cointegration testing, as proposed by Vidyamurthy (2004). Huck & Afawubo identify pairs to test by matching stocks with buy-and-hold returns that are within 10% of the other, over the course of the formation period. Then Cointegration tests are run to find pairs with the highest trace statistic, following Johansen (1988). Differently, Rad, Low and Faff find potential pairs by using the DM, the top 20 pairs with the lowest sum of square deviations, and then testing the lowest SSD pairs eliminating the ones that don’t show evidence of Cointegration following the Engle-Granger (1987) two-step approach. This method of screening cointegrated pairs is the exact same as the DM selection of pairs for trading. This is a selection bias and explains why the results for Cointegration are always below that of the DM. If five of the 20 pairs are found not to be cointegrated, following the method, then the next pairs tested will be pairs that would be ranked 20-25 by the DM. If all these are cointegrated, the returns, in the best-case scenario, will be less than those obtained by the DM. This is a systematic error. Therefore, the comparison of the profitability of the methods is invalid for interpretation. Another comment on the validity of the results, as noted by Huck & Afawubo (2015); is Cointegration-based pairs trading exhibits sensitivity to the width of the opening trigger. Therefore, the true profitability of Cointegration cannot be assessed without parameterizing the opening threshold. Since this is the largest study conducted on pairs trading there is still a gap in the literature for broad research and a need to repeat this analysis using a different pair screening method for the Cointegration pair creation approach. Despite this, a key takeaway can be made about the Distance Method. After transaction costs were considered, monthly excess returns halved indicating vulnerability of pairs trading profitability to such costs, this is consistent of prior work (Do and Faff, 2012) and calls for further research to optimize Pairs Trading for robust profits.

Concluding remarks on the Literature

In general, the literature acknowledges the potential for Pairs trading to yield significant profits when applied correctly. Rather, the focus of discussion revolves around the consistency of this profit and its broader profitability as a long-term trading strategy. The latter being difficult to draw conclusions on due to the variability of the methodology which has been subject to alteration in most comparison research. The literature examines two optimization focuses: processes of pair creation and trading strategy. The former is extensively researched, encompassing the original Distance Method and the Cointegration method inspired by Nobel Laureates Engle and Granger. Studies such as Huck & Afawubo, Caldeira & Moura and Blazquez suggest Cointegration is the most efficient method for pair creation, however the validity of these studies are constrained to the refined samples studied. Further research is required to fill the literature gap of a comprehensive end-to-end comparison between the two methods. The second form of optimization is trade strategy optimization. While studies by Huck (2013) and Huck and Afawubo (2015) display significant positive impact on returns through varying opening thresholds and the length of formation period, there is a lack of formalization of these trading rules. Gulati, Lin, and McCrae (2006) introduced threshold trading rules; a minimum profit per trade, however the rule is flawed, setting this in absolute terms when in the context of pairs trading a more suitable approach would be proportionate, scaling alongside initial investment. Similarly, Caldeira Moura (2013) imposed threshold trading rules by employing stop loss constraints. The positive results of this paper raised questions about potential mitigation of negative results in past papers such as GGR (2006) and Do and Faff (2010), identifying a gap for research to define optimal trading rules.

One such optimization which has not been proposed in the literature is speed of adjustment-based optimization. Tests for Cointegration are in majority conducted adhering to Engle & Granger’s (1987) procedure; this procedure has a notable flaw. Estimation of the long-run equilibrium regression requires choosing one variable (stock in this case) to be dependent in the equation to gather residuals for the first part of the testing procedure. The results of the test are sensitive to the ordering of the regression (Enders, 2014). This can produce misleading results and in the case of pair creation, spurious pairs.

One way around this is using vectors in the specification of the regression equation. Johansen (1988) does this and produces a Cointegration testing method that is robust to this potential error, this is used in the methodology of; Qazi, Rahman, and Gul (2015), Caldeira and Moura (2013), Huck and Afawubo (2015) and Nair (2021). Granger CWJ and Weiss’s (1983) representation theorem allows this to be seen in a form where the speed of adjustment to disequilibrium can be quantified. This may be useful to address some of the issues causing loss of Pairs trading profitability; pairs remaining apart for too long increasing short costs, pairs converging too frequently increasing trading costs and finally the losses from pairs not converging. Since there is no work utilizing this coefficient as a criterion for pairs trading, this is a new area in the literature, one this dissertation evaluates. The empirical investigation of this hypothesis follows.

Methodology

In accordance with the general methodology outlined in the literature, the process of obtaining results has been divided into two phases: pair creation over a ‘formation period’ and subsequent trading of these pairs during a ‘trading period’.

Data Collection

The trading strategy was back tested using daily S&P 500 Stock Close Price data taken from Yahoo finance over the period of 2017-2024. The choice of this subset follows from the suggestions of Huck and Afawubo (2015), who noted that “these stocks are among the most liquid in the world leading to low transaction costs”. This characteristic is significant, as noted by GGR (1999) ‘larger players (institutions) are able to execute trades cheaply’. Therefore, in assessing the profitability of a trading strategy intended for institutional use, it is important to minimize transaction costs. In the GGR and Do and Faff seminal studies, data has been collected on a much larger set of stocks, the entire US equity market. These studies filter stocks if they have ‘one or more days without trading’, with the aim of identifying liquid stocks. By carrying out research on the selected subset of stocks, liquidity has already been achieved.

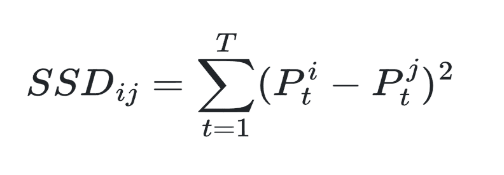
This selected time frame allows assessment of the viability of pairs trading in today’s market, 20 years after the publication of the seminal paper by GGR. Considering the volatility of the global market in recent years the chosen period also provides a good assessment of the market neutrality of pairs trading.

Formation period

The first part of the methodology is the matching of stocks over the formation period, for which two methods of pair matching have been employed: the Distance Method and the Cointegration Method.

The Distance Method

The Distance methodology of pair creation is in accordance with the GGR methodology where stock prices are normalized to $1 on the initial day of the formation period. The Sum of Squared Differences (SSD) of daily normalized stock prices across the formation period is calculated exhaustively for every possible pair combination of stocks. Suitable pairs for trading are then identified based on minimizing SSD. The formula for calculating the SSD follows:



All pairs are ranked according to their SSD, the top 5, top 20, and top 20 after 100 pairs (101-120) are selected for trading, in total 40 pairs of stocks are chosen for trading.

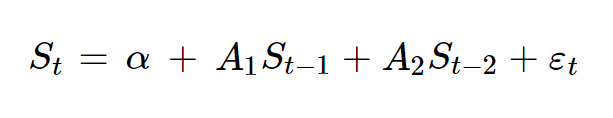
The Cointegration Method

Following from Engle and Granger’s concept of Cointegration (1987), a non-stationary Stock price time series is defined to be integrated of order one when the first difference of the price series is stationary. Two Stock Price series with the same order of integration, are said to be cointegrated when the linear combination of the two Stocks Price series is an order of integration one less than that of the individual prices. The Stocks are defined to be cointegrated and bound together by a long run equilibrium relationship.

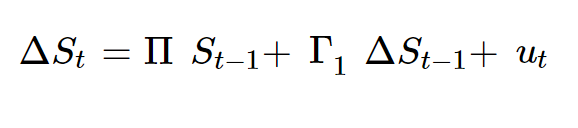
To test for Cointegration, the initial step involves pretesting the order of integration of each individual Stock’s normalized price series, and filtering the data set down to include only Stock’s which are integrated of order one. In the presence of cointegration, the linear combination of two stocks will be stationary. The property of cointegration, which corrects errors or disequilibrium, allows deviations from equilibrium to be traded against, making cointegrated stocks theoretically suitable candidates for pairs trading. To implement the first step, two successive Augmented Dickey Fuller tests are carried out; the first on the Stock’s normalized price series to determine if a Stock is non-stationary and secondly on the first difference of each non-stationary Stock’s price series to determine whether the stock is integrated of order one. The stocks are then retained for cointegration testing.

Cointegration amongst pairs is tested following the Johansen (1988) procedure. The selection of this method of testing was driven by several considerations. The Engle-Granger method, an alternative approach to cointegration testing is sensitive to the ordering of variables in the long run equilibrium regression. Whilst this sensitivity is removed with an infinitely large sample, due to asymptotic theory, which makes the residuals of any variable arrangement regression equal (Enders, 2014); this method of cointegration testing is unsuitable for the pair creation process over a finite formation period. Using the Engle Granger method increases the risk of retaining spurious, non-cointegrated pairs.

Moreover, determining which stock is a dependent variable and which is independent poses a challenge. Since the Johansen test is carried out on a Vector error correction model (VECM) a reparameterization of a Vector Autoregressive model (VAR) which treats each variable symmetrically; this decision does not have to be made and the potential for spurious results is removed. The Vector autoregressive model, considering an example lag order of 2 is depicted below where represents a vector containing two Stocks individual price time series. Here denotes the first lag and the second lag. s a constant and represents the residual term.



The reparameterization of the VAR; the VECM, derived through differencing, is used for testing of cointegrating relationships. The rank of the coefficient matrix of the error correction term , determines the presence of a cointegrated long-run equilibrium relationship. Since pairs of stocks are being assessed, a single co-integrating relationship is tested. The resulting VECM derived from the above VAR of order 2 model is as follows:

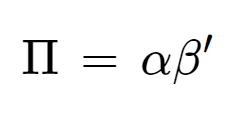


In this methodology, Johansen co-integration tests are exhaustively conducted on all possible pair combinations. Whilst there was contention in the literature on the feasibility of this and there is common practice of pair screening prior to cointegration testing, due to the relatively small number of stocks sampled in this research (500 vs entire US equity market GGR) exhaustive testing is feasible. The test statistic used for the Johansen test is the trace statistic and a restricted constant has been specified. The choice to include a constant is important as omitting a large constant could significantly distort the estimation of other parameters, which are critical for spread calculations and the research question. Additionally, another critical consideration for obtaining valid results is selection of the correct number of lags in the underlying VAR. As empirically shown by Emerson (2007) ‘the number of cointegrating relationships can be very sensitive to lag order chosen for the Underlying VAR’. Equally important, an incorrect number of lags can return inaccurate coefficients of the Pi matrix and subsequent factorized matrices. To address this problem, the Schwarz information criteria has been used to obtain the optimal lag length for the VAR.

One of the key assumptions for estimation of a VAR is that , a vector of the individual processes error terms, is a white noise disturbance. Should the error terms exhibit serial autocorrelation, the OLS estimates for coefficients would be invalidated. To prevent spurious results, a Breusch-Godfrey test is carried out on the VAR of each pair prior to cointegration testing. Pairs which have no residual autocorrelation are carried into cointegration testing, and those whose VAR exhibits residual autocorrelation present are filtered. The preference for the Breusch-Godfrey test over the Durbin-Watson test is driven by its capacity for testing of higher-order pair VAR models.

Speed of Adjustment Coefficient Filtering

When two variables are cointegrated the coefficient matrix of the error correction term in the VECM can be factorized into two matrixes the ‘loading matrix’ or ‘speed of adjustment matrix’ () and the matrix of the cointegrating vectors (), a matrix of the weights placed on each variable that makes the linear combination stationary.



The ‘loading matrix’ is the focus of this thesis’s research question. The loading matrix contains ‘speed of adjustment ‘coefficients which measure the rate at which each variable adjusts to disequilibrium per period.

In response to the critique of the DM by Do and Faff, who attributed the decline in profitability of pairs trading to the corresponding increase in non – converging pairs. This research proposes the use of the speed of adjustment coefficients as a filter mechanism to refine candidate cointegrated pairs for trading. The rationale behind this is that through selection of pairs for trading that have high speed of adjustment to error from the long run cointegrating relationship; this targeted selection can potentially mitigate the problem of losses associated with pairs that fail to converge. Moreover, filtering based on this criterion has the potential to increase profitability over a fixed trading period through facilitating more trades thanks to the favourable speed of adjustment properties of a ‘high speed’ pair. Furthermore, the use of this coefficient offers practical cost-cutting advantages for pairs trading. Given that one leg of the trade is taking up a short position, having high speed pairs can help reduce short interest costs, thereby improving net returns after consideration of short interest payments.

A notable non-quantifiable benefit of using pairs with high speed of adjustment coefficients is the increased confidence in reconvergence. Traders may perceive trading with such pairs as less risky due to the increased assurance of reconvergence. Furthermore, given the speed of adjustment coefficient quantifies the adjustment to disequilibrium over each period, this statistic could also be used to calculate optimal position holding periods.

A close-up of a number

Description automatically generated

To investigate the effect of the proposed filter, the two speed of adjustments coefficients () for each cointegrated pair, which quantify each individual stocks rate of adjustment to disequilibrium/divergence from the long run equilibrium relationship, have been extracted from the VECM’s ‘loading matrix’. These values have then been averaged and stored as the Mean Speed of Adjustment of the cointegrated pair.

To apply the filter, a threshold Mean Speed of Adjustment will be set. Cointegrated pairs from the original pool of eligible pairs, deemed suitable by the cointegration method over the formation period, will be retained if they have a Mean Speed of Adjustment value higher than the threshold and stored into a separate data frame for trading.

To evaluate the correlation between the Mean Speed of Adjustment of pairs and Profitability of the traded pairs, the threshold value used for filtering traded pairs has been systematically adjusted, returning five subsets of pairs, corresponding to six different thresholds mean speed of adjustment values: 0.01, 0.025, 0.05, 0.10, 0.15, 0.20. These values have been selected through analysing the Mean Speed of Adjustment data for all initial cointegrated stocks.

Spread Analysis and the importance of a Formation period.

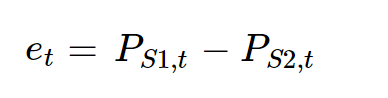
A notable reason to distinguish between a formation period and trading periods, other than for allowing time for traders to analyse and find suitable pairs of stocks for trading, is for creation of trading rules. The trader needs to create opening and closing thresholds to determine when to initiate and exit trading positions. The formation period enables calculation of the standard deviation for the price distance between the traded stocks. This calculation serves as a way of setting a trading rule triggered over the following trading period. For every formation period in the methodology the price distances/spreads between eligible stocks are analysed and the mean and standard deviations of the formation period are used to z-score normalize the price difference/spread between pairs during the trading period. Opening and closing triggers can then be established based on how many standard deviations above or below the mean price distance the traded pairs adjusted price distance is, which is indicated by the z-score, calculated through the Spread Analysis of the formation period.

The second part of the methodology concerns back testing the trading strategy over the trading period, this involves forming trading rules and recording returns.

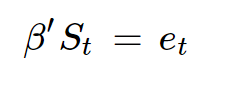
Trading Period

After obtaining sets of pairs of stocks suitable for pairs trading over the formation period, using each pair creation methodology, trading commences the following day. Since trading is based upon the normalized price distance between stocks, trading positions are dependent on the z-score of the price distance observed on a given day within the trading period. As previously mentioned, positions are initiated and exited depending on a threshold standard deviation trigger. In alignment with existing literature methodologies the opening threshold value will be varied across two standard deviations values; 2 and 3. This approach aims to obtain a complete assessment of the returns of each pair formation strategy and importantly for the research question, to account for the sensitivity of cointegration-based pairs trading noted by Huck and Afawubo (2015).

For the DM the ‘spread’ is the nominal price distance which is obtained simply by subtracting the normalized price of Stock 2 from that of Stock 1 in the pair.



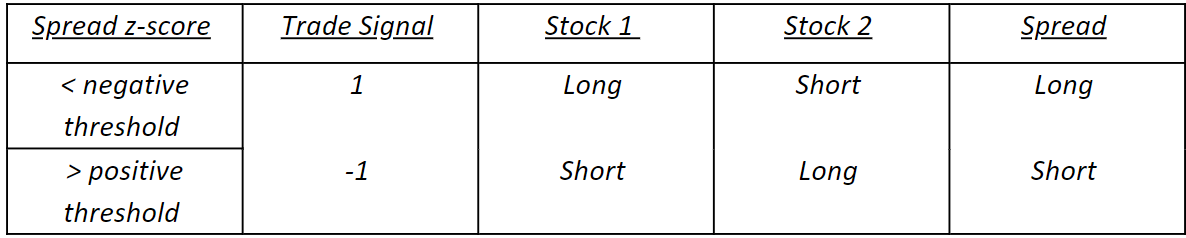
Alternatively, in the cointegration method the ‘spread’ is the stationary series , also known as the cointegration error (Rad et al, 2015). Any non-zero value is the size of the deviation/error from the long run cointegrated equilibrium.



Taking the case of 2 standard deviations as the opening threshold:

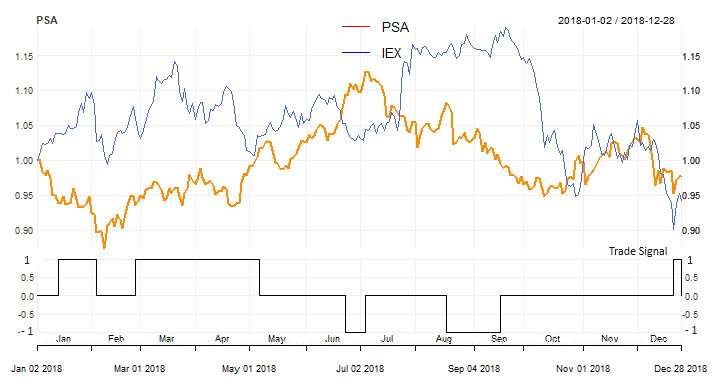
If the z-score of the spread exceeds 2 this indicates that the spread is positive and wide. Stock 1 is a higher price than Stock 2 (or the weighted price in the cointegration case). A short position is initiated for Stock 1 and a long position for Stock 2, this is shorting the spread in effect, denoted by a trade signal of –1. It is expected that Stock 1 will drop in price and Stock 2 will rise, leading to convergence of the spread and a decrease in the z score towards 0.

Conversely, if the z-score of the spread is below -2, this indicates the spread is negative and wide, Stock 1 is a lower price than Stock 2 (or the weighted price). A long position is established on Stock 1 and a short position for Stock 2. A long position is taken on the spread, denoted by a trade signal of 1. The expected outcome is for Stock 1 to increase in price while Stock 2 falls, there will be convergence of the spread and the z-score of the spread will rise towards 0.



Revisiting the original PSA IEX example, in Figure 2 the trade signals have been illustrated alongside the stocks time series. Long and short positions can be seen to be opened in response to fluctuation in the visible spread between the stocks.

Figure 2



Once a trading position on the spread is initiated, it will remain open until the spread crosses the closing threshold. When this occurs, the long position stock will be sold, and the short position stock will be bought. For this research the closing threshold is set at zero. While Caldeira and Moura introduced varying closing thresholds for short and long positions, acknowledging the rationality behind this, results in this paper will be computed holding a constant closing threshold. The reason for this is to mirror the Do and Faff methodology and to focus on assessment of the impact of speed of adjustment filtering to address the poor returns of the cointegration method.

On the final day of the trading period, any pairs with open positions are closed and returns are computed.

Stop Loss Constraint

A Stop Loss constraint has been incorporated into the trading strategy, aligning with the intuition provided by Caldeira and Moura (2013) that ‘stop losses are fundamental in practice to avoid large losses. This incorporation aims to mitigate the risk that large, unconstrained losses can distort the calculation of average returns.

Under the stop loss constraint, positions are closed if the z-score normalized spread continues to diverge past the opening threshold for trading and surpasses the predefined stop loss threshold of the spread. Once the spread surpasses the stop loss threshold, trading for the pair is suspended until the spread returns to below the original opening level.

To obtain a suitable dataset for analysis of the trading strategy, three versions of the same dataset have been generated: one without any stop loss constraint, one with a 25% divergence threshold, and another with a 50% divergence threshold.

Returns Computation

The seminal papers examine two measures of ‘excess returns’ with the term ‘excess’ appropriate as returns in the literature denote opening of long-short positions equal to a dollar. In this study, returns are calculated as decimal-formatted percentages on a stock position of volume one in each leg of the trade, ensuring the same interpretation.

The two computed measures are ‘excess returns on committed capital’ and ‘excess returns on actual employed capital’. The former divides total payoffs by the number of pairs selected for trading, the latter divides total payoffs by the number of pairs that open during the trading period. (GGR). The reason for calculation of the former is to ‘account for the opportunity cost to hedge funds of having to commit capital to a strategy even if it does not trade’ (GGR, 1999). For this research, comparison of pair creation methods’ profitability will be investigated using the ‘actual employed capital’ excess return calculation. The rationale behind this, in line with other design decisions, is to fulfil the aim of comparing the excess returns of a set of pairs traded with a speed of adjustment filter against a set of pairs traded without filtering. However, it's recognized that return on committed capital is a more conservative measure that mimics what a hedge fund might use to report returns (Rad et al, 2015).

Individual returns per trade are calculated by determining the decimal-formatted percentage difference between the price of a stock when a position is opened and the stock price when the position is closed. The percentage difference for each leg of the pair trade is then summated to calculate total return on that individual trade. ‘Total returns’ are calculated by adding every return per trade that occurs within the trading period. This aggregate is then divided by the number of pairs traded to calculate excess returns on actual employed capital. To align with the conventional returns format in the literature, the returns have been expressed in monthly form through dividing by the number of months in the trading period. Additionally, for further insight, the mean return per trade is calculated. This metric accounts the number of trades made per pair. It is calculated by dividing the total returns of a pair by the number of trades and averaging this value across all traded pairs.

Returns are also calculated in the seminal papers following a ‘delayed trading rule’ Do and Faff (2010) whereby positions are opened and closed a day after the trigger. Due to the prevalence of high-frequency trading, with Nasdaq estimating 50 percent of stock trading volume being driven by computer-backed frequency trading (Nasdaq, 2021), this set of returns has not been calculated. This decision has been made because incorporating this consideration, in computing results in this manner, would not be realistic given the advancements in trading technology since the seminal papers.

Length of Formation and Trading Periods

While the paper of Huck and Afawubo (2015) considers two different lengths for the formation period, this research design follows the approach of GGR (1999) and Do and Faff (2010) where pairs are matched over a constant 12-month formation period.

This research design differs from the general literature methodology by introducing variation of the length of the trading period, 6 and 12 months. This is done to investigate the main cause of losses, non-convergent pairs. By varying the length of the trading period an assessment can be made on whether these losses are attributed to; a systematic error in the methodology of not allowing sufficient time for pairs to reconverge, or if they result from weaknesses in the pair creation methods. Given Huck and Afawubo find that this non-convergent behaviour represents approximately 50% of all pairs, it is important to investigate this key observation of the trading strategy.

Assessment of the hypothesis that faster adjusting pairs yield higher employed capital returns should not be impacted by introducing trading period length as a variable. This is because a longer trading period allows more trades to occur, potentially providing more opportunities to yield returns from faster adjusting pairs.

Transaction costs

Transaction cost considerations are built upon the framework established by Rad et al (2015), who employ Dynamic Transaction Costs over varying periods of time. Their approach is appropriate for their analysis given their use of data spanning a large timeframe from 1962 to 2009. For this research design, as the period analysed is relatively short, transaction costs are held constant. Furthermore, to ensure currency, up to date transaction cost data has been collected.

Using data from Interactive Brokers (2024a), the commission per trade is 35 basis points (0.0035). Since each complete trade comprises of four transactions (two per leg, long and short), the total commission paid per trade amounts to 140 basis points (0.0140) or 1.4%. To implement this, 1.4% will be deducted from the returns computed for each pairs trade, produce results Net of transaction costs. Regarding Short sale costs, Interactive Brokers (2024b) impose no short sale cost for transactions up to $100,000. Given a quantity of only one stock is bought of any company and no stock exceeds $100,000 in the S&P 500, BRK-A has been excluded due to its price, this means that short sale costs do not need to be factored in the calculation of returns.

Although one hypothesized advantage of trading with high speed of adjustment cointegrated pairs was to minimize short interest costs, these costs are considered null in this methodology. Nevertheless, for institutional trading involving large stock quantities this remains a significant benefit to consider.

Results

The initial part of the result analysis applies the pairs trading strategy to 2003-2009, the timeframe identified in Do and Faff that characterized diminishing returns. The approach mirrors that of Do and Faff with the primary difference being the sample of stocks used; S&P 500 stocks instead of the entire US Equity market. Additionally, a stop loss of 1.5 times the z-score normalized spread has been implemented.

Consistent with Do and Faff, the Trading Period length for analysis of this period has been narrowed to 6 months. Table 1 reports the results of the analysis. All tables referenced are henceforth located in the appendix. The Monthly Excess Returns on Actual Employed Capital show the weakness in profits observed by Do and Faff for the DM of Pair creation. The results observe monthly excess returns of 11.5 basis points (bps), for a Standard Deviation (SD) opening threshold of 2. This aligns with the 11 bps returns reported by Do and Faff over the same period of analysis. The consistency of these findings with Do and Faff 2010 validates the results in this analysis of the timeframe and in analysis of other time periods going forward.

The conclusions regarding the profitability of the Cointegration Method over this period vary depending on the opening threshold of Standard Deviation (SD), as reported in Table 2. For the Cointegration method without Speed of Adjustment filtering, the monthly excess returns for SD opening thresholds 2 and 3 are found to be 3.2 bps and 10.2 bps, respectively. The observed variance in results, between the two thresholds, aligns with the reporting's of Huck and Afawubo (2015) who observe that the Cointegration method is significantly more profitable when using a 3 SD opening threshold compared to a 2 SD opening threshold. However, despite these results being positive, the returns translate to weak annual returns gross of tax, amounting to 38.4 bps (0.384% annually) and 122.4 bps (1.224% annually).

Analysing the impact of the Speed of Adjustment filter, Table 2 presents the results for varying thresholds of Speed of Adjustment filtered pairs. For an SD opening of 2, only the first level of filtering (0.01 speed of adjustment) has any positive impact on excess returns. Moreover, there appears to be a negative correlation between the speed of adjustment threshold and returns per pair. As the level of filtering increases, it corresponds to a decrease in profits generated and eventually leads to an increase in the losses incurred at this SD opening.

For the 3 SD opening threshold, the relevant SD for co-integration following the findings of Huck and Afawubo, the opposite is observed. For each increment in the degree of Speed of Adjustment filtering, there is a corresponding increase in the returns of the trading strategy. Notably, the highest monthly excess returns yielded by the Cointegration method were with a speed of adjustment threshold value of 0.2, the highest degree of filtering. This indicates a positive correlation between the mean speed of adjustment coefficient per pair and returns over the analysed period, at this SD opening.

While these findings differ from the largely positive findings of Huck and Afawubo over the same period, this inconsistency can be attributed to the difference in strategy design whereby in Huck and Afawubo’s design, for non-traded pairs $2 is put into a long position of the market index. In addition, Huck and Afawubo use rolling formation periods starting every month, leading to six overlapping portfolios of eligible pairs at any one time. As opposed to this study which uses the same single portfolio of pairs from the original formation period, which may lead to comparatively lower monthly returns. Whilst returns are computed effectively per pair in this study, variation in the different selection of pairs traded at any one time may result in cumulative monthly returns.

The Cointegration method is then further examined through the mean return per trade metric, which factors the number of trades made per pair.

As illustrated in Table 2, similar conclusions are drawn: for an SD of 2, a negative correlation is observed between returns and the speed of adjustment of the traded pairs. Whereas, for an SD of 3, a positive correlation is observed. Notably, the subset of filtered pairs with highest mean return per trade comprises of pairs with the highest speed of adjustment (0.2 threshold), which yield a 1.2% return per trade.

Across all variations of pair creation, Distance and Cointegration, and across both values of the SD trading rule, once transaction costs are considered, the trading strategy incurs a loss, supporting the conclusion of Do and Faff regarding the weakness of the strategy. While Speed of Adjustment filtering appears to have a positive effect on returns for the Cointegration Method with a SD opening of 3, the increase in excess returns is not enough to make profits robust to transaction costs. Furthermore, the returns generated are considerably low, a rational investor would rather invest in a risk-free asset, which would yield better returns.

A further analysis of pair behaviour for each pair creation method is presented in Table 3. Across both SD opening triggers, the Average number of days open per trade generally decreases with a higher speed of adjustment threshold. This trend supports the underlying coefficients’ meaning and suggests that a stricter speed of adjustment threshold correlates with shorter trade durations.

Addressing the critique of the DM by Do and Faff (2010) and the suggested reason for lack of profits, non-convergent pairs, the proportion of non-convergent pairs to total pairs traded is recorded. Analysis reveals that filtering based on speed of adjustment doesn’t significantly impact the proportion of non-convergent pairs, as evidenced by general stability in the proportion of non-convergent pairs across the increments of cointegration speed of adjustment filtering. However, there is a clear difference in the average proportion of non-convergent pairs between the two SD opening threshold groups, with the SD 2 results displaying a lower proportion of non-convergent pairs compared to the SD 3 results.

Analysing the proportion of loss-making pairs, there is no clear trend based on the Cointegration Method threshold, although it can be observed that the average proportion of loss-making pairs is lower on average for the Cointegration method than for the DM for both SD result groups. Notably, for the SD 3 results group across all methods, the proportion of loss-making pairs is higher than for the SD 2 results group.

The results of the two discussed metrics contradict the suggestions of Do and Faff that non-convergence was the cause of falling profits. Interestingly, as the proportion of non-convergent pairs increases, the proportion of loss-making pairs decreases, as shown by the contrast in averages across the two SD result groups. These finding challenges Do and Faff’s (2010) concluding remarks that ‘profitable pairs trading is as much about identifying and excluding divergent pairs as it is about identifying convergent pairs’.

The apparent inverse relationship between the proportion of non-convergent pairs and the proportion of loss-making pairs opposed the proposed correlation between profitability and the proportion of non-convergent pairs. While it would be naive to conclude that profits increase with the number of non-convergent pairs, it can be inferred that the proportion of non-convergent pairs does not contribute as significantly to the profitability of the trading strategy as estimated.

In studies such as Rad et al (2015), where the majority of pairs are found to be loss-making and only the minority are profitable, consistent with the findings of this analysis, it could be argued that the profitability of the trading strategy is more influenced by the size of profits rather than the number of losses, which is further emphasized by the stop loss constraint imposed, which mitigates the size of losses.

One suggestion to possibly address this is to consider setting the opening threshold even higher. While it’s observed that the proportion of pairs traded with respect to the number of pairs generated decreases from ~ 95% to 85% when the opening threshold is increased from 2 to 3 SD, this may lead to higher profits. Since profits are determined directly from the amount of convergence, with a higher threshold the size of profits made by profitable trades increases, when the closing threshold is held constant at zero.

Given that analysis has found that an increase in non-converging pairs does not have a significant impact on the proportion of loss-making pairs there is no harm to this, pairs that aren’t traded do not result in losses and those that are, may win larger increasing the profitability of the trading strategy. With the risk of losses from further diverging pairs covered by the stop loss function which can be optimized depending on the initial opening threshold i.e., tighter stop loss constraints on a higher opening SD threshold.

Turning to the second period of analysis 2017-2024, with aim to assess the trading strategy in recent years and reinforce the evaluation of the impact of Speed of Adjustment Filtering on the strategy's returns.

Three stop loss thresholds were applied to calculate returns for this period. The decision to implement these thresholds followed observation of very large losses for the average returns of various methods of pair creation. After looking into returns on individual pairs and observing losses larger than 200% for certain pairs, it became evident that this was clearly a weakness in the methodology. As discussed in the research design, in practice traders will limit losses and close positions at threshold levels of loss. The exclusion of this in the methodology of the seminal papers, GGR (1999) and Do and Faff (2010), is a weakness of their research design. Consequently, two further sets of results were computed for stop losses of 1.25 and 1.5 times the opening threshold. A third set of results using a higher critical value of the spread, was computed to investigate if pairs were being closed prematurely, constraining profits. Through increasing the threshold traded pairs were allowed room for variation. As discussed earlier, for higher thresholds a tighter stop loss constraint may be optimal, while a looser stop loss constraint may be more suitable for lower thresholds. This can be tailored according to factors such as: the size of the positions taken, the spread value at which positions are entered and market conditions. For the analysis of the second time period the focus is on results obtained using a stop loss threshold filter of 1.5, while the monthly returns for no stop loss and a stop loss of 1.25 can be seen in Tables 4 and 5.

Continuing with the structure of analysis of the Do and Faff period the Monthly Gross and Net Excess Returns can be seen in Tables 6 and 7. Despite the implementation of stop loss thresholds, the results indicate consistent negative returns across nearly all pair creation methods. The only profitable pair creation method was the 3 SD opening DM, returning 13 bps monthly returns (1.56 % annually) for the 6-month trading period and 8 bps (0.96 % annually) for the 12-month trading period. These weak gross returns were insufficient to offset transaction costs, resulting in overall losses. No variation of the cointegration method of pair creation was profitable, net of transaction costs, and a comparison can only be made on determining which degree of speed of adjustment filtering incurred the least loss, following the same analysis approach as in the previous period discussed.

Tables 8 and 9 break down the returns associated with each level of Cointegration filtering. For all indicators, there is a negative correlation between the Speed of Adjustment Coefficient of the traded pairs and the profitability measure. This trend can be seen across both opening threshold variations and for both trading period lengths. To make sense of these results, which go against the hypothesis, the behaviour of pairs within each filtered cointegration group are analysed. The pair behaviour analysis is presented in Tables 10 and 11 for each trading period length variation.

Across both SD threshold openings and both variations in trading period length, a clear negative correlation between the speed of adjustment coefficient and the average number of days open per trade can be observed, once again providing validity to the coefficient of speed of adjustment obtained and portraying the underlying meaning of the coefficient.

In contrast to the findings of the Do and Faff analysis period, there is a significant decrease in proportion of non-convergent pairs across all four strategy variations as the speed of adjustment filter is incrementally increased. Twice, the speed of adjustment filter of 0.2 returns the lowest proportion of non-convergent pairs for the 6-month 2 SD group and the 12 month 3-SD group. For the other pairs filtered at 0.15 and 0.1 they exhibit proportionally the fewest non-convergent pairs. While the reduction in non-convergence of traded pairs through speed of adjustment filtering is evident, the way in which this translates to profitability of the strategy is subject for further consideration. Filtering cointegrated pairs by the speed of adjustment coefficient results in a filtered group with, on average, half the proportion of non-convergent pairs compared to the DM in the 2 SD case, and two- thirds of the proportion in the 3 SD case.

As outlined in the methodology, inclusion of the trade period length as a variable aimed to view whether the number of non-convergent pairs was systematically influenced by not allowing pairs enough time to reconverge. The observed proportion of non-convergent pairs is consistently lower for the 12- month trading period design compared to the 6-month period across all groups and SD thresholds, suggesting that this statistic is influenced by the length of the trading period.

On analysis of the proportion of loss-making pairs it is difficult to draw conclusions from observation. Therefore, to gain a more thorough assessment of the relationship between the proportion of non-convergent pairs and the proportion of loss-making pairs, a Linear regression of both metrics has been conducted which can be viewed in Figure 3. The regression uses data spanning all variations of results obtained from different standard deviations and trading period lengths. The regression confirms the previously observed negative relationship between the proportion of pair non-convergence and the proportion of loss-making pairs with a regression coefficient of –0.29. This relationship is statistically significant, indicated by the associated p-value. Furthermore, the robustness of the regression results has been confirmed through a Breusch-Pagan test for residual heteroscedasticity, alongside a visual examination of the regression residuals, shown in Figure 4. The implication of the weak negative relationship, indicated by the low magnitude coefficient, is that the suggested reason for loss-making pairs, as argued by Do and Faff, is not accounted to pair non-convergence.

This conclusion is supported by the weak R-squared value of the regression analysis, leading to the notion that other factors besides pair non-convergence contribute significantly to the occurrence of loss-making pairs. A logical explanation for this is the way in which a pair is identified as non-convergent, as this is simply a measure of the number of pairs that are left open at the end of the trading period. This approach introduces bias towards pairs that are opened later in the trading period, potentially overlooking pairs that were closed due to stop loss and incurred losses.

To better understand the effect of pair behaviour on profitability, a metric such as the average distance converged by a pair could provide more insight. This metric would measure the average distance converged by a pair offering a better understanding of the causes of profits, and in particular a better understanding to assess the impact of speed of adjustment filtering.

Whilst speed of adjustment filtering has demonstrated a positive influence on the number of convergent pairs and the behaviour of pairs; reflected in consistently lower average days traded per pair, there are other factors, such as opening trigger thresholds and the degree of stop loss, that appear to have a more significant impact on the profitability of the trading strategy.

Based on this conclusion, an additional set of results have been generated using a reduced sample of stocks, specifically the first 100 in the S&P 500 index. To assess the conclusions of the prior analysis, the new analysis involves varying the standard deviation opening threshold across six new values while implementing a stop loss set at 1.5 times the z-score normalized spread. The final set of results are summarized in Table 12. These results underline the significant impact the opening threshold has on the strategy’s returns, showing differing returns for different opening triggers for the same pool of filtered pairs. Through varying the opening threshold, previous negative returns of selected pair groups have turned positive. Notable groups included the Cointegration Method Filtered Groups of 0.1 and 0.15, where gross monthly excess returns have changed from – 10 bps (- 1.2% annually) and –118 bps (- 13.2% annually), to robust positive returns of 36.4 bps (4.368% annually) and 27.4 bps (3.28% annually) respectively. Whilst it is imperative to acknowledge that these findings are based on a small sample of stocks rather than the complete selection used in prior analysis, these results highlight the importance of selecting the appropriate opening threshold to maximize profits.

Given that only one stock was generated for the following subset, the results should be interpreted critically. However, it’s noteworthy that the highest returns were observed for the Cointegration Method filtered Threshold 0.2, with a SD opening trigger of 4. Remarkably, the single pair yielded monthly gross excess returns of 163 bps (19% annually) and monthly net returns of 128 bps (15.3% annually).

Upon examination of the standard deviation thresholds that are most profitable for each subset of the trading strategy, highlighted in red in table 12, it appears that lower opening thresholds are most profitable for pairs with a lower mean speed of adjustment, while higher opening thresholds yield the largest returns for pairs with higher speed of adjustment. This suggests the potential for further research to evaluate optimal trading thresholds based on the characteristics of pairs traded.

The prior concluded negative relationship between pair non-convergence and the mean speed of adjustment remains consistent. It is notable however, that the proportion of non-convergent pairs tends to increase with the magnitude of the opening threshold. This is intuitive as pairs with larger opening thresholds are likely to remain further apart. Given this, it seems theoretically sound that high opening thresholds and fast-converging pairs, which return low proportions of non-convergent pairs, could be a compatible combination for a profit maximizing trading strategy.

The final remark for these results is that the stop loss used is held constant at 1.5. As discussed earlier, it goes well that for higher opening thresholds a tighter stop loss would be beneficial, and this is not a parameter that should be held constant for a set of results but should be adjusted alongside the opening threshold and pair creation method in order to optimize the trading strategy.

Conclusion

The trend identified in the seminal papers regarding the decline in pairs trading profits has continued into recent years. Across both prominent pair creation methods, the Cointegration and Distance methods, pairs trading continues to yield losses before transaction costs and further losses net of transaction costs, in the modern market. In response to the proposed cause of losses suggested by Do and Faff; pair non-convergence, this study proposed a solution to non-convergence through filtering cointegrated pairs based on the mean speed of adjustment value for the pair. To investigate this hypothesis, analysis was carried out over the original period of weak returns identified by Do and Faff from 2003-2009 and the speed of adjustment filter was back tested.

The returns remained unaffected by the implementation of a speed of adjustment filter, over this period, and despite observing a decrease in the average days traded per pair with the filter, this did not correspond to a reduction in the proportion of non-convergent pairs or a resulting decrease in the proportion of loss-making pairs, as these metrics remained consistent across all degrees of speed filtering.

The analysis was then extended to the previous 7-year period from 2017-2024, during which the weak returns shown in the Do and Faff period had turned into losses. In the second analysis, conclusions were drawn regarding the causes of loss in the trading strategy and the impact of cointegration-based filtering. As anticipated, introducing a variation in the length of the trading period to 12 months resulted in a reduction in the high proportion of non-convergent pairs, highlighting this metric as a systematically weak indicator when used over a short period of time. While the metric should show a more reliable explanation when measured over a 12-month period, this study advocates for a more suitable measure of pair behaviour: the average convergence per traded pair. Nonetheless, the impact of the speed of adjustment filter remains relevant from observation of the metric. As pairs were filtered by adjustment speed, groups with higher adjustment speed exhibited lower proportions of non-convergent pairs. The use of 7 magnitudes of speed of adjustment filtering allowing for the negative relationship between pair non-convergence and the speed of adjustment per pair to be characterized. The problem of pair non- convergence identified by Do and Faff can be reduced through filtering of pairs based on their speed of adjustment filter.

From observation of the Do and Faff period, the results suggested a negative relationship between the proportion of non-convergent pairs and the proportion of loss-making pairs. To investigate this further a linear regression between the two metrics was conducted on the results of the modern period of analysis, the regression concluding that there was indeed a negative relationship between the two metrics. This suggests that the factors influencing trading losses extend beyond non-convergence, as supported by the low fit of the regression.

Considering this, with aim of underpinning the determinants of the trading strategies profits, the opening thresholds were assessed, testing the statement of Huck and Afawubo that ‘Pairs trading returns exhibit sensitivity to the width of the opening trigger’. A further dataset was obtained for which the standard deviation SD opening threshold was varied six times. The findings revealed significant variation in profitability assessment based on the opening threshold, providing insight into the critical importance of selecting the appropriate opening threshold. Given the optimal opening threshold varied with the speed of adjustment of the pairs traded, the opening threshold should be finetuned to suit the characteristics of the pairs traded. The implementation of speed of adjustment filtering facilitates this, with fast adjusting pairs proving compatible with higher opening thresholds, resulting in substantial profits over the sampled data.

Final Remarks

The adoption of Speed of Adjustment filtering has addressed the metric of pair non convergence raised by Do and Faff (2010). However, through analysis it has been found that this metric alone does not significantly contribute to the strategy’s profitability. Instead, variables such as the opening trigger and stop loss demonstrate greater influence on the return. The consistent characteristic observed in pairs filtered by speed of adjustment offers potential for tailoring the trading strategy. To explore this further, additional research on larger samples and different applications, such as high-frequency pairs trading and pairs trading in cryptocurrencies or security derivatives, is needed to test the integration of this metric and optimize the profits of the trading strategy.

Appendix

|  |  |
| --- | --- |
| Abbreviation Table | |
| GGR | Gatev, Goetzmann and Rouwenhorst |
| DM | Distance Model |
|  |  |

Table 1

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| *Period of Analysis* | *2003-01-01 / 2009-12-31* | | | | | | | | | | | | | | | |
| *Trading Period Length* | *6 months* | | | | | | | | | | | | | | | |
| *Pair Creation Method* | *Distance* | | *Cointegration* | | | | | | | | | | | | | |
| *Speed of Adjustment Filtering Threshold (Cointegration)* | *NA* | | *0* | | *0.01* | | *0.025* | | *0.05* | | *0.1* | | *0.15* | | *0.2* | |
| *Opening Threshold (Standard Deviation)* | *2* | *3* | *2* | *3* | *2* | *3* | *2* | *3* | *2* | *3* | *2* | *3* | *2* | *3* | *2* | *3* |
| *Monthly Excess Returns on Actual Employed Capital* | *0.00115* | *-0.00017* | *0.00032* | *0.00102* | *0.00034* | *0.00105* | *0.00031* | *0.00129* | *0.00001* | *0.00159* | *-0.00036* | *0.00190* | *-0.00400* | *0.00111* | *-0.00698* | *0.00300* |
| *Monthly Net Excess Returns on Actual Employed Capital* | *-0.00357* | *-0.00357* | *-0.00470* | *-0.00273* | *-0.00473* | *-0.00271* | *-0.00489* | *-0.00253* | *-0.00549* | *-0.00240* | *-0.00604* | *-0.00232* | *-0.00935* | *-0.00299* | *-0.01212* | *-0.00124* |

Table 2

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| *Standard Deviation* | *Cointegration Speed of Adjustment Filtering Threshold* | *Monthly Excess Returns on Actual Employed Capital* | *Monthly Net Excess Returns on Actual Employed Capital* | *Mean Return Per Trade* | *Mean Net Return Per Trade* |
| *2* | *0* | *0.00032* | *-0.00470* | *-0.002086586* | *-0.016086586* |
| *0.01* | *0.00034* | *-0.00473* | *-0.002554817* | *-0.016554817* |
| *0.025* | *0.00031* | *-0.00489* | *-0.003441124* | *-0.017441124* |
| *0.05* | *0.00001* | *-0.00549* | *-0.005431589* | *-0.019431589* |
| *0.1* | *-0.00036* | *-0.00604* | *-0.007817727* | *-0.021817727* |
| *0.15* | *-0.00400* | *-0.00935* | *-0.025922841* | *-0.039922841* |
| *0.2* | *-0.00698* | *-0.01212* | *-0.036990902* | *-0.050990902* |
| *3* | *0* | *0.00102* | *-0.00273* | *0.004125079* | *-0.009874921* |
| *0.01* | *0.00105* | *-0.00271* | *0.004281932* | *-0.009718068* |
| *0.025* | *0.00129* | *-0.00240* | *0.004484142* | *-0.009515858* |
| *0.05* | *0.00159* | *-0.00240* | *0.004918589* | *-0.009081411* |
| *0.1* | *0.00190* | *-0.00232* | *0.005057197* | *-0.008942803* |
| *0.15* | *0.00111* | *-0.00299* | *0.005453832* | *-0.008546168* |
| *0.2* | *0.00300* | *-0.00124* | *0.012809006* | *-0.001190994* |

Table 3

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| *Opening Threshold (Standard Deviation)* | *Pair Creation Method* | *Average number of days open per trade* | *Proportion of Non-Convergent Pairs* | *Proportion of Loss-making Pairs* |
| *2* | *Cointegration Method Threshold 0* | *16.82173888* | *0.221344853* | *0.657172298* |
| *Cointegration Method Threshold 0.01* | *16.05247813* | *0.21643552* | *0.663811838* |
| *Cointegration Method Threshold 0.025* | *15.18487487* | *0.208537033* | *0.671796222* |
| *Cointegration Method Threshold 0.05* | *13.75790812* | *0.201465303* | *0.688036019* |
| *Cointegration Method Threshold 0.1* | *13.00575946* | *0.202474558* | *0.669181227* |
| *Cointegration Method Threshold 0.15* | *13.52555983* | *0.208409748* | *0.612055766* |
| *Cointegration Method Threshold 0.2* | *14.1384485* | *0.179191292* | *0.629688731* |
| *Distance Method* | *18.67516995* | *0.274965654* | *0.708521325* |
| *3* | *Cointegration Method Threshold 0* | *34.80302953* | *0.352815456* | *0.59635382* |
| *Cointegration Method Threshold 0.01* | *34.15386904* | *0.345992706* | *0.599798005* |
| *Cointegration Method Threshold 0.025* | *32.73109702* | *0.334057168* | *0.605712123* |
| *Cointegration Method Threshold 0.05* | *29.40084031* | *0.325275868* | *0.605212077* |
| *Cointegration Method Threshold 0.1* | *26.44766199* | *0.349842644* | *0.588293446* |
| *Cointegration Method Threshold 0.15* | *27.93178223* | *0.349178347* | *0.58882524* |
| *Cointegration Method Threshold 0.2* | *29.88533534* | *0.401488989* | *0.613826326* |
| *Distance Method* | *28.75690574* | *0.388351913* | *0.657129586* |

Table 4

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| No Stop Loss Threshold | | | | |
| *Trading Period Length* | *Method Name* | *SD* | *Monthly Excess Returns on Actual Employed Capital* | *Monthly Net Excess Returns on Actual Employed Capital* |
| *6 months* | *Cointegration Method Threshold 0* | *2* | *-0.002287781* | *-0.005320461* |
| *6 months* | *Cointegration Method Threshold 0* | *3* | *-0.001814032* | *-0.004431779* |
| *6 months* | *Cointegration Method Threshold 0.01* | *2* | *-0.002809384* | *-0.005888387* |
| *6 months* | *Cointegration Method Threshold 0.01* | *3* | *-0.00225006* | *-0.004892259* |
| *6 months* | *Cointegration Method Threshold 0.025* | *2* | *-0.002820969* | *-0.005983767* |
| *6 months* | *Cointegration Method Threshold 0.025* | *3* | *-0.002251314* | *-0.004937495* |
| *6 months* | *Cointegration Method Threshold 0.05* | *2* | *-0.004069364* | *-0.007454967* |
| *6 months* | *Cointegration Method Threshold 0.05* | *3* | *-0.003248983* | *-0.006072086* |
| *6 months* | *Cointegration Method Threshold 0.1* | *2* | *-0.002175725* | *-0.006208231* |
| *6 months* | *Cointegration Method Threshold 0.1* | *3* | *-0.000859022* | *-0.004161066* |
| *6 months* | *Cointegration Method Threshold 0.15* | *2* | *-0.016580003* | *-0.020570081* |
| *6 months* | *Cointegration Method Threshold 0.15* | *3* | *-0.013625974* | *-0.016723674* |
| *6 months* | *Cointegration Method Threshold 0.2* | *2* | *-0.026719258* | *-0.03043138* |
| *6 months* | *Cointegration Method Threshold 0.2* | *3* | *-0.019451744* | *-0.022416896* |
| *6 months* | *Distance Method* | *2* | *-0.000162959* | *-0.003808211* |
| *6 months* | *Distance Method* | *3* | *0.000830744* | *-0.002095336* |
| *12 months* | *Cointegration Method Threshold 0* | *2* | *-0.001030468* | *-0.002940327* |
| *12 months* | *Cointegration Method Threshold 0* | *3* | *-0.000699323* | *-0.002244018* |
| *12 months* | *Cointegration Method Threshold 0.01* | *2* | *-0.00141808* | *-0.003372468* |
| *12 months* | *Cointegration Method Threshold 0.01* | *3* | *-0.001025378* | *-0.002598386* |
| *12 months* | *Cointegration Method Threshold 0.025* | *2* | *-0.001864427* | *-0.003892332* |
| *12 months* | *Cointegration Method Threshold 0.025* | *3* | *-0.001448173* | *-0.003067751* |
| *12 months* | *Cointegration Method Threshold 0.05* | *2* | *-0.003201212* | *-0.005414167* |
| *12 months* | *Cointegration Method Threshold 0.05* | *3* | *-0.002647826* | *-0.004386875* |
| *12 months* | *Cointegration Method Threshold 0.1* | *2* | *-0.005907045* | *-0.008527581* |
| *12 months* | *Cointegration Method Threshold 0.1* | *3* | *-0.005079492* | *-0.00711967* |
| *12 months* | *Cointegration Method Threshold 0.15* | *2* | *-0.031214552* | *-0.033883518* |
| *12 months* | *Cointegration Method Threshold 0.15* | *3* | *-0.029574206* | *-0.031603585* |
| *12 months* | *Cointegration Method Threshold 0.2* | *2* | *-0.049974712* | *-0.052628289* |
| *12 months* | *Cointegration Method Threshold 0.2* | *3* | *-0.044570597* | *-0.046667061* |
| *12 months* | *Distance Method* | *2* | *-0.000575148* | *-0.002971426* |
| *12 months* | *Distance Method* | *3* | *0.000281715* | *-0.001487201* |

Table 5

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Stop Loss Threshold of 1.25 | | | | |
| *Trading Period Length* | *Method Name* | *SD* | *Monthly Excess Returns on Actual Employed Capital* | *Monthly Net Excess Returns on Actual Employed Capital* |
| *6 months* | *Cointegration Method Threshold 0* | *2* | *-0.000636989* | *-0.005481195* |
| *6 months* | *Cointegration Method Threshold 0* | *3* | *-0.001067778* | *-0.005152228* |
| *6 months* | *Cointegration Method Threshold 0.01* | *2* | *-0.000667543* | *-0.005621789* |
| *6 months* | *Cointegration Method Threshold 0.01* | *3* | *-0.001036798* | *-0.005184556* |
| *6 months* | *Cointegration Method Threshold 0.025* | *2* | *-0.000538507* | *-0.005694514* |
| *6 months* | *Cointegration Method Threshold 0.025* | *3* | *-0.000687285* | *-0.004968231* |
| *6 months* | *Cointegration Method Threshold 0.05* | *2* | *-0.000410589* | *-0.006013261* |
| *6 months* | *Cointegration Method Threshold 0.05* | *3* | *-0.00030068* | *-0.004866547* |
| *6 months* | *Cointegration Method Threshold 0.1* | *2* | *-0.001087074* | *-0.007740131* |
| *6 months* | *Cointegration Method Threshold 0.1* | *3* | *-0.001790565* | *-0.007375858* |
| *6 months* | *Cointegration Method Threshold 0.15* | *2* | *-0.003461219* | *-0.009923692* |
| *6 months* | *Cointegration Method Threshold 0.15* | *3* | *-0.008849936* | *-0.01450118* |
| *6 months* | *Cointegration Method Threshold 0.2* | *2* | *-0.004508315* | *-0.010426497* |
| *6 months* | *Cointegration Method Threshold 0.2* | *3* | *-0.016651766* | *-0.022737625* |
| *6 months* | *Distance Method* | *2* | *-0.000971861* | *-0.007406975* |
| *6 months* | *Distance Method* | *3* | *0.000942679* | *-0.003780206* |
| *12 months* | *Cointegration Method Threshold 0* | *2* | *-0.000316129* | *-0.003883881* |
| *12 months* | *Cointegration Method Threshold 0* | *3* | *-0.000288334* | *-0.003152858* |
| *12 months* | *Cointegration Method Threshold 0.01* | *2* | *-0.000329882* | *-0.003998901* |
| *12 months* | *Cointegration Method Threshold 0.01* | *3* | *-0.000313629* | *-0.003258335* |
| *12 months* | *Cointegration Method Threshold 0.025* | *2* | *-0.000330954* | *-0.004163351* |
| *12 months* | *Cointegration Method Threshold 0.025* | *3* | *-0.00023669* | *-0.003304177* |
| *12 months* | *Cointegration Method Threshold 0.05* | *2* | *-0.000308557* | *-0.004508929* |
| *12 months* | *Cointegration Method Threshold 0.05* | *3* | *-0.000112101* | *-0.003422528* |
| *12 months* | *Cointegration Method Threshold 0.1* | *2* | *-0.000620254* | *-0.005464471* |
| *12 months* | *Cointegration Method Threshold 0.1* | *3* | *-0.001001644* | *-0.004889374* |
| *12 months* | *Cointegration Method Threshold 0.15* | *2* | *-0.001568419* | *-0.006360054* |
| *12 months* | *Cointegration Method Threshold 0.15* | *3* | *-0.007177199* | *-0.011079765* |
| *12 months* | *Cointegration Method Threshold 0.2* | *2* | *-0.002353071* | *-0.007283122* |
| *12 months* | *Cointegration Method Threshold 0.2* | *3* | *-0.013630641* | *-0.017767299* |
| *12 months* | *Distance Method* | *2* | *-0.000335995* | *-0.005313648* |
| *12 months* | *Distance Method* | *3* | *0.000473125* | *-0.002873762* |

Table 6

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| *Period of Analysis* | *2017-01-01 / 2023-12-31* | | | | | | | | | | | | | | | |
| *Trading Period Length* | *6 months* | | | | | | | | | | | | | | | |
| *Pair Creation Method* | *Distance* | | *Cointegration* | | | | | | | | | | | | | |
| *Speed of Adjustment Filtering Threshold (Cointegration)* | *NA* | | *0* | | *0.01* | | *0.025* | | *0.05* | | *0.1* | | *0.15* | | *0.2* | |
| *Opening Threshold (Standard Deviation)* | *2* | *3* | *2* | *3* | *2* | *3* | *2* | *3* | *2* | *3* | *2* | *3* | *2* | *3* | *2* | *3* |
| *Monthly Excess Returns on Actual Employed Capital* | *-0.000591995* | *0.001392246* | *-0.000833339* | *-0.001455191* | *-0.000890747* | *-0.001536862* | *-0.00072549* | *-0.001049192* | *-0.000695231* | *-0.00043139* | *-0.001757274* | *-0.000943046* | *-0.004784059* | *-0.011070083* | *-0.007940994* | *-0.022964753* |
| *Monthly Net Excess Returns on Actual Employed Capital* | *-0.005901242* | *-0.002368767* | *-0.004982519* | *-0.004794152* | *-0.005134379* | *-0.004929088* | *-0.005139959* | *-0.004536267* | *-0.005498366* | *-0.004171984* | *-0.007629378* | *-0.0055532* | *-0.0103653* | *-0.015640574* | *-0.013190993* | *-0.027914753* |

Table 7

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| *Period of Analysis* | *2017-01-01 / 2023-12-31* | | | | | | | | | | | | | | | |
| *Trading Period Length* | *12 months* | | | | | | | | | | | | | | | |
| *Pair Creation Method* | *Distance* | | *Cointegration* | | | | | | | | | | | | | |
| *Speed of Adjustment Filtering Threshold (Cointegration)* | *NA* | | *0* | | *0.01* | | *0.025* | | *0.05* | | *0.1* | | *0.15* | | *0.2* | |
| *Opening Threshold (Standard Deviation)* | *2* | *3* | *2* | *3* | *2* | *3* | *2* | *3* | *2* | *3* | *2* | *3* | *2* | *3* | *2* | *3* |
| *Monthly Excess Returns on Actual Employed Capital* | *-0.000311242* | *0.000860485* | *-0.000395589* | *-0.000440649* | *-0.000437999* | *-0.000510058* | *-0.000432196* | *-0.000416208* | *-0.000455584* | *0.000148191* | *-0.001070973* | *-0.001020747* | *-0.003503637* | *-0.011879947* | *-0.00607937* | *-0.019461941* |
| *Monthly Net Excess Returns on Actual Employed Capital* | *-0.004341851* | *-0.001720142* | *-0.003373036* | *-0.002706944* | *-0.00350203* | *-0.002840936* | *-0.003636505* | *-0.002840598* | *-0.003978878* | *-0.002787447* | *-0.005243093* | *-0.004155341* | *-0.007535269* | *-0.014930333* | *-0.010223982* | *-0.022796074* |

Table 8

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| *2017-01-01 / 2023-12-31* | | | | | |
| *6 Month Trading Period* | | | | | |
| *Standard Deviation* | *Cointegration Speed of Adjustment Filtering Threshold* | *Monthly Excess Returns on Actual Employed Capital* | *Monthly Net Excess Returns on Actual Employed Capital* | *Mean Return Per Trade* | *Mean Net Return Per Trade* |
| *2* | *0* | *-0.000833339* | *-0.004982519* | *-0.005484811* | *-0.019484811* |
| *0.01* | *-0.000890747* | *-0.005134379* | *-0.005918093* | *-0.019918093* |
| *0.025* | *-0.000725489* | *-0.005139959* | *-0.005176584* | *-0.019176584* |
| *0.05* | *-0.000695231* | *-0.005498366* | *-0.004718789* | *-0.018718789* |
| *0.1* | *-0.001757274* | *-0.007629378* | *-0.006245757* | *-0.020245757* |
| *0.15* | *-0.004784059* | *-0.0103653* | *-0.016520167* | *-0.030520167* |
| *0.2* | *-0.007940993* | *-0.013190993* | *-0.030431414* | *-0.044431414* |
| *3* | *0* | *-0.001455191* | *-0.004794152* | *-0.00757957* | *-0.02157957* |
| *0.01* | *-0.001536862* | *-0.004929088* | *-0.007911257* | *-0.021911257* |
| *0.025* | *-0.001049191* | *-0.004536267* | *-0.005490619* | *-0.019490619* |
| *0.05* | *-0.00043139* | *-0.004171984* | *-0.003093481* | *-0.017093481* |
| *0.1* | *-0.000943046* | *-0.0055532* | *-0.007170205* | *-0.021170205* |
| *0.15* | *-0.011070083* | *-0.015640574* | *-0.042113113* | *-0.056113113* |
| *0.2* | *-0.022964753* | *-0.027914753* | *-0.052047551* | *-0.066047551* |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
| Table 9 |  |  |  |  |  |
| *2017-01-01 / 2023-12-31* | | | | | |
| *12 Month Trading Period* | | | | | |
| *Standard Deviation* | *Cointegration Speed of Adjustment Filtering Threshold* | *Monthly Excess Returns on Actual Employed Capital* | *Monthly Net Excess Returns on Actual Employed Capital* | *Mean Return Per Trade* | *Mean Net Return Per Trade* |
| *2* | *0* | *-0.000395589* | *-0.003373036* | *-0.005357529* | *-0.019357529* |
| *0.01* | *-0.000437999* | *-0.00350203* | *-0.00594145* | *-0.01994145* |
| *0.025* | *-0.000432196* | *-0.003636505* | *-0.005856571* | *-0.019856571* |
| *0.05* | *-0.000455584* | *-0.003978878* | *-0.005661217* | *-0.019661217* |
| *0.1* | *-0.001070973* | *-0.005243093* | *-0.007438494* | *-0.021438494* |
| *0.15* | *-0.003503637* | *-0.007535269* | *-0.020083048* | *-0.034083048* |
| *0.2* | *-0.00607937* | *-0.010223982* | *-0.031763626* | *-0.045763626* |
| *3* | *0* | *-0.000440649* | *-0.002706944* | *-0.00491027* | *-0.01891027* |
| *0.01* | *-0.000510058* | *-0.002840936* | *-0.006020993* | *-0.020020993* |
| *0.025* | *-0.000416208* | *-0.002840598* | *-0.006043041* | *-0.020043041* |
| *0.05* | *-0.000148191* | *-0.002787447* | *-0.005045078* | *-0.019045078* |
| *0.1* | *-0.001020747* | *-0.004155341* | *-0.012865024* | *-0.026865024* |
| *0.15* | *-0.011879947* | *-0.014930333* | *-0.103944801* | *-0.117944801* |
| *0.2* | *-0.019461941* | *-0.022796074* | *-0.135571322* | *-0.149571322* |

Table 10

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 2017-01-01 / 2023-12-31 | | | | |
| 6 Month Trading Period | | | | |
| *Opening Threshold (Standard Deviation)* | *Pair Creation Method* | *Average number of days open per trade* | *Proportion of Non-Convergent Pairs* | *Proportion of Loss-making Pairs* |
| *2* | *Cointegration Method Threshold 0* | 14.69299599 | 0.170336628 | 0.686541939 |
| *Cointegration Method Threshold 0.01* | 13.59206988 | 0.157996828 | 0.694922652 |
| *Cointegration Method Threshold 0.025* | 12.47170686 | 0.150831827 | 0.697984077 |
| *Cointegration Method Threshold 0.05* | 11.60571825 | 0.145653771 | 0.693681526 |
| *Cointegration Method Threshold 0.1* | 10.45631109 | 0.121995311 | 0.702456811 |
| *Cointegration Method Threshold 0.15* | 10.52639401 | 0.171523603 | 0.73347868 |
| *Cointegration Method Threshold 0.2* | 10.16474567 | 0.111363636 | 0.579545455 |
| *Distance Method* | 16.57267984 | 0.263176638 | 0.75772792 |
| *3* | *Cointegration Method Threshold 0* | 31.85715686 | 0.280386577 | 0.629027128 |
| *Cointegration Method Threshold 0.01* | 30.6810631 | 0.270026423 | 0.637638519 |
| *Cointegration Method Threshold 0.025* | 28.70942994 | 0.252717101 | 0.642488955 |
| *Cointegration Method Threshold 0.05* | 24.99307251 | 0.225753732 | 0.643075494 |
| *Cointegration Method Threshold 0.1* | 20.56598411 | 0.205743797 | 0.638819609 |
| *Cointegration Method Threshold 0.15* | 20.61413691 | 0.20645967 | 0.648862275 |
| *Cointegration Method Threshold 0.2* | 19.70401154 | 0.237662338 | 0.712554113 |
| *Distance Method* | 26.65888483 | 0.33856722 | 0.643565115 |
|  |  |  |  |  |
|  |  |  |  |  |
| Table 11 |  |  |  |  |
|  |  |  |  |  |
| 2017-01-01 / 2023-12-31 | | | | |
| 12 Month Trading Period | | | | |
| *Opening Threshold (Standard Deviation)* | *Pair Creation Method* | *Average number of days open per trade* | *Proportion of Non-Convergent Pairs* | *Proportion of Loss-making Pairs* |
| *2* | *Cointegration Method Threshold 0* | 18.99766513 | 0.15717996 | 0.683267231 |
| *Cointegration Method Threshold 0.01* | 16.96493681 | 0.146036804 | 0.69309637 |
| *Cointegration Method Threshold 0.025* | 15.02563353 | 0.135132268 | 0.699541925 |
| *Cointegration Method Threshold 0.05* | 13.23437071 | 0.131956977 | 0.70184839 |
| *Cointegration Method Threshold 0.1* | 11.6303424 | 0.105201603 | 0.735623926 |
| *Cointegration Method Threshold 0.15* | 11.30494839 | 0.082682668 | 0.745020136 |
| *Cointegration Method Threshold 0.2* | 10.47534873 | 0.105808081 | 0.653282828 |
| *Distance Method* | 18.50677283 | 0.167521368 | 0.777991453 |
| *3* | *Cointegration Method Threshold 0* | 42.43036996 | 0.245477876 | 0.62068351 |
| *Cointegration Method Threshold 0.01* | 40.01140932 | 0.230056787 | 0.630558061 |
| *Cointegration Method Threshold 0.025* | 35.95018588 | 0.210735847 | 0.637322886 |
| *Cointegration Method Threshold 0.05* | 29.83408843 | 0.195139915 | 0.640944887 |
| *Cointegration Method Threshold 0.1* | 24.09639567 | 0.161337035 | 0.669473324 |
| *Cointegration Method Threshold 0.15* | 25.32095687 | 0.172183472 | 0.742070893 |
| *Cointegration Method Threshold 0.2* | 23.7773569 | 0.142171717 | 0.756060606 |
| *Distance Method* | 35.18270788 | 0.265237353 | 0.664965927 |

Figure 3

A screenshot of a computer program

Description automatically generated

Figure 4

A graph with a line and dots

Description automatically generated

Table 12

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| 2017-01-01 / 2023-12-31 | | | | | | |
| 12 Month Trading Period | | | | | | |
| *Pair Creation Method* | *SD* | *Monthly Excess Returns on Actual Employed Capital* | *Monthly Net Excess Returns on Actual Employed Capital* | *Proportion of Non-Convergent Pairs* | *Proportion of Loss-making Pairs* | *Number of pairs generated* |
| *Cointegration Method Threshold 0* | *2.5* | *0.00062* | *-0.00183* | *0.21* | *0.62* | *512.2* |
| *2.75* | *0.00053* | *-0.00179* | *0.23* | *0.61* | *512.2* |
| *3.25* | *0.00051* | *-0.00156* | *0.28* | *0.59* | *512.2* |
| *3.5* | *0.00010* | *-0.00190* | *0.30* | *0.61* | *512.2* |
| *3.75* | *-0.00019* | *-0.00211* | *0.33* | *0.61* | *512.2* |
| *4* | *-0.00037* | *-0.00221* | *0.35* | *0.61* | *512.2* |
| *Cointegration Method Threshold 0.01* | *2.5* | *0.00064* | *-0.00187* | *0.19* | *0.63* | *439.3* |
| *2.75* | *0.00062* | *-0.00176* | *0.21* | *0.61* | *439.3* |
| *3.25* | *0.00064* | *-0.00150* | *0.27* | *0.60* | *439.3* |
| *3.5* | *0.00018* | *-0.00188* | *0.29* | *0.61* | *439.3* |
| *3.75* | *-0.00018* | *-0.00214* | *0.31* | *0.61* | *439.3* |
| *4* | *-0.00037* | *-0.00225* | *0.34* | *0.61* | *439.3* |
| *Cointegration Method Threshold 0.025* | *2.5* | *0.00040* | *-0.00222* | *0.17* | *0.65* | *323.8* |
| *2.75* | *0.00052* | *-0.00196* | *0.19* | *0.62* | *323.8* |
| *3.25* | *0.00064* | *-0.00159* | *0.24* | *0.60* | *323.8* |
| *3.5* | *0.00008* | *-0.00206* | *0.27* | *0.62* | *323.8* |
| *3.75* | *-0.00014* | *-0.00218* | *0.29* | *0.62* | *323.8* |
| *4* | *-0.00037* | *-0.00231* | *0.31* | *0.62* | *323.8* |
| *Cointegration Method Threshold 0.05* | *2.5* | *0.00074* | *-0.00204* | *0.17* | *0.63* | *150.5* |
| *2.75* | *0.00100* | *-0.00167* | *0.18* | *0.61* | *150.5* |
| *3.25* | *0.00075* | *-0.00165* | *0.23* | *0.60* | *150.5* |
| *3.5* | *0.00035* | *-0.00196* | *0.26* | *0.61* | *150.5* |
| *3.75* | *-0.00021* | *-0.00242* | *0.27* | *0.62* | *150.5* |
| *4* | *-0.00035* | *-0.00246* | *0.29* | *0.62* | *150.5* |
| *Cointegration Method Threshold 0.1* | *2.5* | *-0.00060* | *-0.00355* | *0.10* | *0.67* | *14.3* |
| *2.75* | *0.00149* | *-0.00135* | *0.12* | *0.62* | *14.3* |
| *3.25* | *0.00364* | *0.00113* | *0.17* | *0.54* | *14.3* |
| *3.5* | *0.00286* | *0.00047* | *0.20* | *0.57* | *14.3* |
| *3.75* | *0.00199* | *-0.00029* | *0.21* | *0.65* | *14.3* |
| *4* | *0.00065* | *-0.00156* | *0.21* | *0.57* | *14.3* |
| *Cointegration Method Threshold 0.15* | *2.5* | *-0.00196* | *-0.00522* | *0.07* | *0.82* | *3.7* |
| *2.75* | *-0.00225* | *-0.00505* | *0.00* | *0.69* | *3.7* |
| *3.25* | *-0.00343* | *-0.00577* | *0.00* | *0.82* | *3.7* |
| *3.5* | *-0.00054* | *-0.00233* | *0.00* | *0.76* | *3.7* |
| *3.75* | *0.00165* | *-0.00008* | *0.13* | *0.62* | *3.7* |
| *4* | *0.00274* | *0.00100* | *0.13* | *0.69* | *3.7* |
| *Cointegration Method Threshold 0.2* | *2.5* | *-0.00091* | *-0.00674* | *0.00* | *1.00* | *1.0* |
| *2.75* | *-0.00635* | *-0.00985* | *0.00* | *1.00* | *1.0* |
| *3.25* | *0.00001* | *-0.00349* | *0.00* | *1.00* | *1.0* |
| *3.5* | *0.00955* | *0.00605* | *0.00* | *0.00* | *1.0* |
| *3.75* | *0.00948* | *0.00598* | *0.00* | *0.00* | *1.0* |
| *4* | *0.01631* | *0.01281* | *0.00* | *0.00* | *1.0* |
| *Distance Method* | *2.5* | *0.00012* | *-0.00269* | *0.23* | *0.67* | *40.0* |
| *2.75* | *-0.00004* | *-0.00255* | *0.26* | *0.65* | *40.0* |
| *3.25* | *0.00001* | *-0.00211* | *0.30* | *0.65* | *40.0* |
| *3.5* | *-0.00028* | *-0.00226* | *0.31* | *0.65* | *40.0* |
| *3.75* | *-0.00041* | *-0.00234* | *0.34* | *0.63* | *40.0* |
| *4* | *-0.00066* | *-0.00247* | *0.36* | *0.62* | *40.0* |

Code written to obtain data implemented in RStudio

# Required Packages

library(rvest)

library(quantmod)

library(forecast)

library(tidyverse)

library(vars)

library(tsDyn)

library(urca)

library(lubridate)

library(tidyr)

library(dplyr)

options(scipen = 999, digits = 10)

#-------------------------------------------------------------------------------

# Code for obtaining data

tbl <- read\_html('https://en.wikipedia.org/wiki/List\_of\_S%26P\_500\_companies') %>% html\_nodes(css = 'table')

tbl <- tbl[1] %>% html\_table() %>% as.data.frame()

tickers <- tbl$Symbol

Comnames <- as.data.frame(tbl$Symbol, tbl$Security)

Stock\_information <- data.frame(Symbol = tbl$Symbol, Security = tbl$Security, GICS\_Sector = tbl$GICS.Sector)

quotes <- new.env()

getSymbols(tickers, src = 'yahoo', from = '2017-01-01', env = quotes)

closeprices <- do.call(merge, lapply(quotes, Cl))

colnames(closeprices) <- gsub(".Close", "", colnames(closeprices), fixed = TRUE)

closepricesfull <- closeprices[, colSums(is.na(closeprices))== 0]

write.zoo(closepricesfull, file = "pricedata.csv")

rm(closeprices)

rm(closepricesfull)

prices <- as.xts(read.zoo("C:/Dissertation/pricedata.csv", header = TRUE, index.column = 1))

#-------------------------------------------------------------------------------

# Order of Integration testing

intorder <- function(dataset){

prices <- dataset

adfp <- numeric(length = ncol(prices))

for(i in 1:ncol(prices)){

test <- ur.df(prices[,i], type = c("trend"))

adfp[i] <- (test@teststat[1] < test@cval[1,2])

}

adfp2 <- numeric(length = ncol(prices))

for( i in 1:ncol(prices)){

if(adfp[i] == 0){

v <- na.omit(diff(prices[,i]))

test2 <- ur.df(v, type = c("drift"))

adfp2[i] <- (test2@teststat[1] < test2@cval[1,2])

}

else{}

}

prices[, adfp2 == 1]

}

I1closeprices <- intorder(prices)

rm(prices)

# To optimize computational performance, minimize the number of objects in the environment when running this code.

#-------------------------------------------------------------------------------

# Obtain percentage returns for each date, percentage returns = (difference between day and last day) / last day

percentagedailyreturns <- function(dataset){

percentret <- dataset[-1,]

for (i in 1: ncol(percentret)){

percentret[,i] <- diff(dataset[,i])[-1]/ lag(dataset[,i])[-1]

}

return(percentret)

}

percret <- percentagedailyreturns(I1closeprices)

#-------------------------------------------------------------------------------

#retindex take xts of stock prices and returns an xts of prices normalised to initial value of 1, multiplies previous day stock by (one plus daily percentage returns)

retindex <- function(dataset){

new <- dataset

rets <- percentagedailyreturns(dataset)

rets <- coredata(rets)

newmatrix <- coredata(new)

newmatrix[1,] <- 1

for(i in 2: nrow(new)){

newvals <- numeric(length = ncol(newmatrix))

for(j in 1: ncol(newmatrix)){

newvals[j] <- newmatrix[i-1,j] \* (1 + rets[i-1, j])

}

newmatrix[i,] <- newvals

}

coredata(new) <- newmatrix

return(new)

}

normalisedclosedprices <- retindex(I1closeprices)

#-------------------------------------------------------------------------------

#Spread analysis function

calculate\_spread\_analysis <- function(spread\_xts) {

spread\_analysis <- data.frame(PairSpread = character(),

Mean\_of\_spread = numeric(),

SD\_spread = numeric(),

stringsAsFactors = FALSE)

for (i in 1:ncol(spread\_xts)) {

Mean\_of\_spread <- mean(spread\_xts[, i])

SD\_spread <- sd(spread\_xts[, i])

spread\_analysis <- rbind(spread\_analysis,

data.frame(PairSpread = colnames(spread\_xts)[i],

Mean\_of\_spread = Mean\_of\_spread,

SD\_spread = SD\_spread,

stringsAsFactors = FALSE))

}

return(spread\_analysis)

}

#-------------------------------------------------------------------------------

# Distance Method of Pair Creation

distancemethod <- function(formationperiod, tradingperiod) {

I1closeprices\_formation <- I1closeprices[formationperiod]

I1closeprices\_trading <- I1closeprices[tradingperiod]

normalised\_formation <- retindex(I1closeprices\_formation)

normalised\_trading <- retindex(I1closeprices\_trading)

SSD\_dataframe <- data.frame(Pair = character(), Sum\_of\_Square\_Distance = numeric())

Distance\_CloseSet\_list <- list()

Distance\_spread\_xts <- xts(matrix(NA, nrow = nrow(I1closeprices), ncol = 0), order.by = index(I1closeprices))

for (i in 1:(ncol(I1closeprices)-1)){

for (j in(i+1):ncol(I1closeprices)){

DistanceCloseset <- cbind(normalised\_formation[,i], normalised\_formation[,j])

Distance <- DistanceCloseset[,1] - DistanceCloseset[,2]

Squared\_Distance <- (Distance^2)

Stock1 <- colnames(DistanceCloseset[,1])

Stock2 <- colnames(DistanceCloseset[,2])

DistanceCloseset <- cbind(DistanceCloseset,Distance = Distance, Squared\_Distance = Squared\_Distance)

colnames(DistanceCloseset) <- c(Stock1, Stock2 , "Distance", "Squared\_Distance")

Sum\_of\_Square\_Distance <- sum(Squared\_Distance)

Pair\_name <- paste(Stock1, Stock2, sep = "\_")

Distance\_CloseSet\_list[[paste(Stock1, Stock2, sep = "\_")]] <- DistanceCloseset

SSD\_dataframe <- rbind(SSD\_dataframe, data.frame(Pair = Pair\_name, Sum\_of\_Square\_Distance = Sum\_of\_Square\_Distance))

}

}

SSD\_dataframe\_sorted <- SSD\_dataframe[order(SSD\_dataframe$Sum\_of\_Square\_Distance), ]

top\_5 <- SSD\_dataframe\_sorted[1:5, ]

top\_20 <- SSD\_dataframe\_sorted[1:20, ]

top\_pairs\_100\_to\_200 <- SSD\_dataframe\_sorted[101:120,]

top\_20\_unique <- top\_20[!top\_20$Pair %in% top\_5$Pair, ]

GGR\_Distance\_Pairs <- rbind(top\_5, top\_20\_unique, top\_pairs\_100\_to\_200)

rownames(GGR\_Distance\_Pairs) <- NULL

Distance\_spread\_formation\_xts <- xts(matrix(NA, nrow = nrow(I1closeprices\_formation), ncol = 0), order.by = index(I1closeprices\_formation))

Stocks\_in\_pair <- list()

for(i in 1:nrow(GGR\_Distance\_Pairs)) {

Stocks\_in\_pair[[i]] <- unlist(strsplit(GGR\_Distance\_Pairs[i,"Pair"], "\_"))

Stockname1 <- Stocks\_in\_pair[[i]][1]

Stockname2 <- Stocks\_in\_pair[[i]][2]

DAsset1 <- normalised\_formation[, Stockname1]

DAsset2 <- normalised\_formation[, Stockname2]

distance\_pair\_name <- paste(Stockname1, Stockname2, "spread", sep = "\_")

Dspread\_formation <- DAsset1 - DAsset2

Distance\_spread\_formation\_xts <- cbind(Distance\_spread\_formation\_xts, Dspread\_formation)

colnames(Distance\_spread\_formation\_xts)[ncol(Distance\_spread\_formation\_xts)] <- paste(Stockname1, Stockname2, "spread", sep = "\_")

}

Distance\_spread\_trading\_xts <- xts(matrix(NA, nrow = nrow(I1closeprices\_trading), ncol = 0), order.by = index(I1closeprices\_trading))

for(i in 1:nrow(GGR\_Distance\_Pairs)) {

Stocks\_in\_pair[[i]] <- unlist(strsplit(GGR\_Distance\_Pairs[i,"Pair"], "\_"))

Stockname1 <- Stocks\_in\_pair[[i]][1]

Stockname2 <- Stocks\_in\_pair[[i]][2]

DAsset1 <- normalised\_trading[, Stockname1]

DAsset2 <- normalised\_trading[, Stockname2]

distance\_pair\_name <- paste(Stockname1, Stockname2, "spread", sep = "\_")

Dspread\_trading <- DAsset1 - DAsset2

Distance\_spread\_trading\_xts <- cbind(Distance\_spread\_trading\_xts, Dspread\_trading)

colnames(Distance\_spread\_trading\_xts)[ncol(Distance\_spread\_trading\_xts)] <- paste(Stockname1, Stockname2, "spread", sep = "\_")

}

spread\_analysis\_formation\_distance <- calculate\_spread\_analysis(Distance\_spread\_formation\_xts)

spread\_analysis\_trading\_distance <- calculate\_spread\_analysis(Distance\_spread\_trading\_xts)

result <- list(SSD\_dataframe = SSD\_dataframe,

Distance\_CloseSet\_list = Distance\_CloseSet\_list,

GGR\_Distance\_Pairs = GGR\_Distance\_Pairs,

I1closeprices\_formation = I1closeprices\_formation,

I1closeprices\_trading = I1closeprices\_trading,

normalised\_formation = normalised\_formation,

normalised\_trading = normalised\_trading,

Distance\_spread\_formation\_xts = Distance\_spread\_formation\_xts,

Distance\_spread\_trading\_xts = Distance\_spread\_trading\_xts,

Spread\_Analysis\_Formation\_Distance = spread\_analysis\_formation\_distance,

Spread\_Analysis\_Trading\_Distance = spread\_analysis\_trading\_distance)

}

#-------------------------------------------------------------------------------

# Cointegration Method of Pair Creation incorporating Speed of Adjustment Filtering, with adjustable threshold values

cointegrationspeedfilteringmethod <- function(formationperiod, tradingperiod){

cointegration\_threshold\_list <- list()

I1closeprices\_formation <- I1closeprices[formationperiod]

I1closeprices\_trading <- I1closeprices[tradingperiod]

normalised\_formation <- retindex(I1closeprices\_formation)

normalised\_trading <- retindex(I1closeprices\_trading)

lagselectlist <- list()

pairs <- list()

cbeta\_and\_constant\_df <- data.frame(Pair = character(), cbeta = numeric(), constant = numeric(), stringsAsFactors = FALSE)

cAlpha\_df <- data.frame(Stock1 = character(), Stock2 = character(), cAlphaS1 = numeric(), cAlphaS2 = numeric(), stringsAsFactors = FALSE)

Cointegrated\_spread\_formation\_filtered\_xts <- xts(matrix(NA, nrow = nrow(I1closeprices\_formation), ncol = 0), order.by = index(I1closeprices\_formation))

Cointegrated\_spread\_trading\_filtered\_xts <- xts(matrix(NA, nrow = nrow(I1closeprices\_trading), ncol = 0), order.by = index(I1closeprices\_trading))

filtered\_pairs <- data.frame(Stock1 = character(), Stock2 = character(),cbeta = numeric(), constant = numeric(), cAlphaS1 = numeric(), cAlphaS2 = numeric(), stringsAsFactors = FALSE)

cointegratedpairsnameslist = list()

for (i in 1:(ncol(I1closeprices\_formation)-1)){

for (j in(i+1):(ncol(I1closeprices\_formation))){

Closeset <- cbind(normalised\_formation[,i], normalised\_formation[,j])

lagselect <- VARselect(Closeset, type = "const")

lags <- lagselect[["selection"]][["AIC(n)"]]

lagselectlist <- append(lagselectlist, lags)

tempvar <- VAR(Closeset, p = lags, type = "const")

stest <- serial.test(tempvar, type = "BG")

if (stest$serial$p.value > 0.05){

ctest <- ca.jo(normalised\_formation[, c(i,j)], type = "trace", ecdet = "const",

K = if(lags > 2){lags} else{2}, spec="transitory")

if (ctest@teststat[2] > ctest@cval[2, 2]){

vec\_model <- cajorls(ctest, r=1)

col\_names <- colnames(vec\_model$rlm$model)

cointegrated\_pair\_names <- colnames(vec\_model$rlm$coefficients)

cointegrated\_pairs\_namesclean <- gsub("\\.d$", "", cointegrated\_pair\_names)

pairnamehere <- paste(cointegrated\_pairs\_namesclean[1], cointegrated\_pairs\_namesclean[2], sep = "\_")

pairs[[pairnamehere]] <- list(rlm = vec\_model$rlm, beta = vec\_model$beta)

cAlphaS1 <- vec\_model$rlm$coefficients[1,1]

cAlphaS2 <- vec\_model$rlm$coefficients[1,2]

cbeta <- (vec\_model$beta[2,1])

constant <- (vec\_model$beta[3,1])

meanspeed <- ((abs(cAlphaS1) + abs(cAlphaS2))/2)

filtered\_pairs <- rbind(filtered\_pairs, data.frame(Stock1 = cointegrated\_pairs\_namesclean[1],

Stock2 = cointegrated\_pairs\_namesclean[2],

cbeta = cbeta,

constant = constant,

cAlphaS1 = cAlphaS1,

cAlphaS2 = cAlphaS2,

meanspeed = meanspeed,

stringsAsFactors = FALSE))

}

}

}

}

speedthresholds <- c(0, 0.01, 0.025 , 0.05, 0.10, 0.15, 0.20)

for (speedthreshold in speedthresholds) {

tryCatch({

filtered\_pairs\_speed <- filtered\_pairs[ filtered\_pairs$meanspeed > speedthreshold, ]

if (!is.null(filtered\_pairs\_speed)){

cointegration\_filtered\_tests <- list()

Asset1 <- filtered\_pairs\_speed$Stock1

Asset2 <- filtered\_pairs\_speed$Stock2

Asset1\_unique <- unique(Asset1)

Asset2\_unique <- unique(Asset2)

all\_assets <- unique(c(Asset1\_unique, Asset2\_unique))

I1closeprices\_formation\_filtered\_all\_assets <- I1closeprices\_formation[, all\_assets]

I1closeprices\_trading\_filtered\_all\_assets <- I1closeprices\_trading[, all\_assets]

normalised\_formation\_filtered\_all\_assets <- normalised\_formation[, all\_assets]

normalised\_trading\_filtered\_all\_assets <- normalised\_trading[, all\_assets]

Cointegrated\_spread\_formation\_filtered\_xts <- xts(matrix(NA, nrow = nrow(normalised\_formation), ncol = nrow(filtered\_pairs\_speed)), order.by = index(normalised\_formation))

Cointegrated\_spread\_trading\_filtered\_xts <- xts(matrix(NA, nrow = nrow(normalised\_trading), ncol = nrow(filtered\_pairs\_speed)), order.by = index(normalised\_trading))

for(k in 1:nrow(filtered\_pairs\_speed)){

Asset1\_k <- as.character(Asset1[k])

Asset2\_k <- as.character(Asset2[k])

cbeta <- filtered\_pairs\_speed$cbeta[k]

A1CL <- normalised\_formation\_filtered\_all\_assets[, Asset1\_k]

A2CL <- normalised\_formation\_filtered\_all\_assets[, Asset2\_k]

constant <- filtered\_pairs\_speed$constant[k]

Spread\_formation <- constant + A1CL + (cbeta \* A2CL)

colnames(Spread\_formation) <- paste(Asset1\_k, Asset2\_k, "spread", sep = "\_")

Cointegrated\_spread\_formation\_filtered\_xts[, k] <- Spread\_formation

A1CLt <- normalised\_trading\_filtered\_all\_assets[,Asset1\_k]

A2CLt <- normalised\_trading\_filtered\_all\_assets[,Asset2\_k]

Spread\_trading <- constant + A1CLt + (cbeta \* A2CLt)

colnames(Spread\_trading) <- paste(Asset1\_k, Asset2\_k, "spread", sep = "\_")

Cointegrated\_spread\_trading\_filtered\_xts[, k] <- Spread\_trading

pairname <- paste(Asset1\_k, Asset2\_k, sep = "\_")

if (pairname %in% names(pairs)) {

cointegration\_filtered\_tests[[pairname]] <- pairs[[pairname]]

}

}

colnames(Cointegrated\_spread\_formation\_filtered\_xts) <- paste(Asset1, Asset2, "spread", sep = "\_")

colnames(Cointegrated\_spread\_trading\_filtered\_xts) <- paste(Asset1, Asset2, "spread", sep = "\_")

spread\_analysis\_formation\_filtered\_cointegration <- calculate\_spread\_analysis(Cointegrated\_spread\_formation\_filtered\_xts)

spread\_analysis\_trading\_filtered\_cointegration <- calculate\_spread\_analysis(Cointegrated\_spread\_trading\_filtered\_xts)

cointegration\_threshold\_list[[paste("Cointegration Method Threshold", speedthreshold)]] <- list(

cointegratedfilteredpairs = filtered\_pairs\_speed,

Cointegrationtests = cointegration\_filtered\_tests,

I1closeprices\_formation = I1closeprices\_formation\_filtered\_all\_assets,

I1closeprices\_trading = I1closeprices\_trading\_filtered\_all\_assets,

normalised\_formation = normalised\_formation\_filtered\_all\_assets,

normalised\_trading = normalised\_trading\_filtered\_all\_assets,

Cointegrated\_spread\_formation\_filtered\_xts = Cointegrated\_spread\_formation\_filtered\_xts,

Cointegrated\_spread\_trading\_filtered\_xts = Cointegrated\_spread\_trading\_filtered\_xts,

spread\_analysis\_formation\_filtered\_cointegration = spread\_analysis\_formation\_filtered\_cointegration,

spread\_analysis\_trading\_filtered\_cointegration = spread\_analysis\_trading\_filtered\_cointegration)

}

},error = function(e){

message <- paste("No pairs above the speed threshold", speedthreshold, "for filtered cointegration")

cointegration\_threshold\_list[[paste("Cointegration Method Threshold", speedthreshold)]]<<- list(message)

})

}

return(cointegration\_threshold\_list)

}

#-------------------------------------------------------------------------------

# Signals and Returns function

# Stop loss threshold has been incorporated, and its value can be adjusted as needed.

generate\_signals\_and\_returns <- function(spread\_analysis\_data, spread\_xts, close\_prices, standard\_deviation\_threshold) {

stoploss\_threshold <- 1.5 \* standard\_deviation\_threshold

input\_name <- deparse(substitute(spread\_analysis\_data))

input\_name\_parts <- strsplit(input\_name, "\_")

suffix <- sapply(input\_name\_parts, function(x) tail(x, n = 1))

spread\_list\_signals\_name <- paste("spread\_with\_nsignal\_",suffix, sep = "")

returns\_list\_name <- paste("returns\_dataframe\_xts\_",suffix, sep = "")

signals\_and\_returns\_list <- list(list(),list())

names(signals\_and\_returns\_list) <- c(spread\_list\_signals\_name, returns\_list\_name)

for (i in 1:nrow(spread\_analysis\_data)){

pairnames <- spread\_analysis\_data[i,1]

pairspreadname <- spread\_analysis\_data$PairSpread[i]

normalisedspread <- ((spread\_xts[, pairspreadname] - spread\_analysis\_data$Mean\_of\_spread[i]) / spread\_analysis\_data$SD\_spread[i])

nsignal <- numeric(nrow(normalisedspread))

stopsignal <- numeric(nrow(normalisedspread))

nsignal[1] <- 0

stopsignal[1] <- 0

if (normalisedspread[1] >= standard\_deviation\_threshold) {

nsignal[1] <- -1

} else if (normalisedspread[1] <= -(standard\_deviation\_threshold)) {

nsignal[1] <- 1

}

for (j in 2:nrow(normalisedspread)) {

if(stopsignal[j-1] == 0) {

if (nsignal[j - 1] == 0) {

if (normalisedspread[j] >= standard\_deviation\_threshold) {

nsignal[j] <- -1

} else if (normalisedspread[j] <= -(standard\_deviation\_threshold)) {

nsignal[j] <- 1

} else {

nsignal[j] <- 0

}

} else if (nsignal[j - 1] == -1) {

if (normalisedspread[j] <= 0) {

nsignal[j] <- 0

} else {

nsignal[j] <- nsignal[j - 1]

}

} else if (nsignal[j - 1] == 1) {

if (normalisedspread[j] >= 0) {

nsignal[j] <- 0

} else {

nsignal[j] <- nsignal[j - 1]

}

}

if (nsignal[j - 1] == -1 && normalisedspread[j] >= stoploss\_threshold) {

nsignal[j] <- 0

stopsignal[j] <- 1

} else if(nsignal[j - 1] == 1 && normalisedspread[j] <= -(stoploss\_threshold)) {

nsignal[j] <- 0

stopsignal[j] <- 1

}

} else {

nsignal[j] <- 0

if (abs(normalisedspread[j]) < standard\_deviation\_threshold) {

stopsignal[j] <- 0

} else {

stopsignal[j] <- 1

}

}

}

nsignal[length(nsignal)] <- 0 # setting the last observation as zero

spread\_with\_nsignal <- xts(cbind(Spread = normalisedspread, nsignal = nsignal, stopsignal = stopsignal), order.by = index(normalisedspread))

signals\_and\_returns\_list[[spread\_list\_signals\_name]][[i]] <- spread\_with\_nsignal

}

for(i in 1:nrow(spread\_analysis\_data)){

pairnames <- spread\_analysis\_data[i,1]

return\_spread <- signals\_and\_returns\_list[[spread\_list\_signals\_name]][[i]][,1]

return\_signal <- signals\_and\_returns\_list[[spread\_list\_signals\_name]][[i]][,2]

stoploss\_signal <- signals\_and\_returns\_list[[spread\_list\_signals\_name]][[i]][,3]

Stock\_names <- strsplit(pairnames, "\_")[[1]]

Stock1\_name <- as.character(Stock\_names[1])

Stock2\_name <- as.character(Stock\_names[2])

Stock1\_ClosePrice <- close\_prices[, Stock1\_name]

Stock2\_ClosePrice <- close\_prices[, Stock2\_name]

returns\_dataframe <- data.frame(

Stock1ClosePrice = Stock1\_ClosePrice,

Stock2ClosePrice = Stock2\_ClosePrice,

Spread = return\_spread,

Signal = return\_signal,

Stoploss\_Signal = stoploss\_signal,

ASSET\_1 = rep(NA, length(return\_spread)),

ASSET\_2 = rep(NA, length(return\_spread)),

Asset\_1\_Points = numeric(nrow(return\_spread)),

Asset\_2\_Points = numeric(nrow(return\_spread)),

PercentageReturnsA1 = rep(NA, length(return\_spread)),

PercentageReturnsA2 = rep(NA, length(return\_spread))

)

colnames(returns\_dataframe)[3:5] <- c("Z-Score Normalized Spread", "Signal", "Stoploss Signal")

returns\_dataframe$ASSET\_1[1] <- Stock1\_ClosePrice[1]

returns\_dataframe$ASSET\_2[1] <- Stock2\_ClosePrice[1]

for(j in 2:nrow(returns\_dataframe)){

if(returns\_dataframe[j, "Signal"] == returns\_dataframe[(j-1), "Signal"]) {

returns\_dataframe[j,"ASSET\_1"] <- returns\_dataframe[j-1,"ASSET\_1"]

} else {

returns\_dataframe[j,"ASSET\_1"] <- returns\_dataframe[j,1]

}

if(returns\_dataframe[j, "Signal"] == returns\_dataframe[(j-1), "Signal"]) {

returns\_dataframe[j,"ASSET\_2"] <- returns\_dataframe[(j-1),"ASSET\_2"]

} else {

returns\_dataframe[j,"ASSET\_2"] <- returns\_dataframe[j,2]

}

}

Asset\_1\_Points <- rep(NA, nrow(returns\_dataframe))

Asset\_2\_Points <- rep(NA, nrow(returns\_dataframe))

for(k in 2:nrow(returns\_dataframe)) {

if(returns\_dataframe[k, "Signal"] != returns\_dataframe[k-1, "Signal"]) {

if(returns\_dataframe[k-1, "Signal"] == 1) {

Asset\_1\_Points[k] <- round(returns\_dataframe[k,"ASSET\_1"] - returns\_dataframe[k-1, "ASSET\_1"], 8)

Asset\_2\_Points[k] <- round(-1 \* (returns\_dataframe[k,"ASSET\_2"] - returns\_dataframe[k-1, "ASSET\_2"]), 8)

}

else if(returns\_dataframe[k-1, "Signal"] == -1) {

Asset\_1\_Points[k] <- round((returns\_dataframe[k,"ASSET\_1"] - returns\_dataframe[k-1, "ASSET\_1"]) \* -1, 8)

Asset\_2\_Points[k] <- round(returns\_dataframe[k,"ASSET\_2"] - returns\_dataframe[k-1, "ASSET\_2"], 8)

}

else if(returns\_dataframe[k-1,"Signal"] == 0){

Asset\_1\_Points[k] <- NA

Asset\_2\_Points[k] <- NA

}

} else {

Asset\_1\_Points[k] <- NA

Asset\_2\_Points[k] <- NA

}

}

returns\_dataframe$Asset\_1\_Points <- Asset\_1\_Points

returns\_dataframe$Asset\_2\_Points <- Asset\_2\_Points

for(p in 2:nrow(returns\_dataframe)){

returns\_dataframe[p, "PercentageReturnsA1"] <- returns\_dataframe[p,"Asset\_1\_Points"]/returns\_dataframe[p-1,"ASSET\_1"]

returns\_dataframe[p, "PercentageReturnsA2"] <- returns\_dataframe[p,"Asset\_2\_Points"]/returns\_dataframe[p-1,"ASSET\_2"]

}

Returns\_dataframe\_xts <- xts(returns\_dataframe, order.by = index(spread\_xts))

signals\_and\_returns\_list[[returns\_list\_name]][[i]] <- Returns\_dataframe\_xts

}

return(signals\_and\_returns\_list)

}

#-------------------------------------------------------------------------------

#Function to Aggregate Returns Across All Pairs

generate\_aggregate\_results <- function(returns\_data\_list\_function) {

Aggregate\_Results <- data.frame(

Pair = character(),

Number\_of\_Trades = integer(),

Total\_Percentage\_Returns = numeric(),

Average\_Percentage\_Returns\_Per\_Trade = numeric(),

Total\_Transaction\_costs = numeric(),

Net\_Total\_Percentage\_Returns = numeric(),

Net\_Average\_Percentage\_Returns\_Per\_Trade = numeric(),

Non\_Convergent\_Pair = integer(),

Total\_Days\_Open = integer(),

Average\_Days\_Open\_Per\_Trade = numeric()

)

result\_list <- list()

for (i in 1:length(returns\_data\_list\_function)) {

Number\_of\_trades <- sum(!is.na(returns\_data\_list\_function[[i]]$Asset\_1\_Points))

if (Number\_of\_trades == 0) {

PairSpread <- paste0(colnames(returns\_data\_list\_function[[i]])[1], "\_", colnames(returns\_data\_list\_function[[i]])[2])

temp\_result <- data.frame(Pair = PairSpread,

Number\_of\_Trades = Number\_of\_trades,

Total\_Percentage\_Returns = NA,

Average\_Percentage\_Returns\_Per\_Trade = NA,

Total\_Transaction\_costs = NA,

Net\_Total\_Percentage\_Returns = NA,

Net\_Average\_Percentage\_Returns\_Per\_Trade = NA,

Non\_Convergent\_Pair = NA,

Total\_Days\_Open = NA,

Average\_Days\_Open\_Per\_Trade = NA)

}else{

Total\_Returnpc <- (sum(returns\_data\_list\_function[[i]]$PercentageReturnsA1, na.rm = TRUE) +

sum(returns\_data\_list\_function[[i]]$PercentageReturnsA2, na.rm = TRUE))

Average\_Returnpc <- Total\_Returnpc / Number\_of\_trades

Transaction\_costs <- Number\_of\_trades \* 0.0140

Net\_Total\_Returnpc <- Total\_Returnpc - Transaction\_costs

Net\_Average\_Returnpc <- Net\_Total\_Returnpc / Number\_of\_trades

PairSpread <- paste0(colnames(returns\_data\_list\_function[[i]])[1], "\_", colnames(returns\_data\_list\_function[[i]])[2])

second\_to\_last\_signal <- returns\_data\_list\_function[[i]]$Signal[nrow(returns\_data\_list\_function[[i]]) - 1]

Non\_Convergent\_Pair <- as.integer(second\_to\_last\_signal != 0)

Total\_Days\_Open <- sum(returns\_data\_list\_function[[i]]$Signal != 0)

Average\_Days\_Open\_Per\_Trade <- Total\_Days\_Open / Number\_of\_trades

temp\_result <- data.frame(Pair = PairSpread,

Number\_of\_Trades = Number\_of\_trades,

Total\_Percentage\_Returns = Total\_Returnpc,

Average\_Percentage\_Returns\_Per\_Trade = Average\_Returnpc,

Total\_Transaction\_costs = Transaction\_costs,

Net\_Total\_Percentage\_Returns = Net\_Total\_Returnpc,

Net\_Average\_Percentage\_Returns\_Per\_Trade = Net\_Average\_Returnpc,

Non\_Convergent\_Pair = Non\_Convergent\_Pair,

Total\_Days\_Open = Total\_Days\_Open,

Average\_Days\_Open\_Per\_Trade = Average\_Days\_Open\_Per\_Trade)

}

result\_list[[i]] <- temp\_result

}

Aggregate\_Results <- do.call(rbind, result\_list)

row.names(Aggregate\_Results) <- NULL

return(Aggregate\_Results)

}

#-------------------------------------------------------------------------------

# Script to Generate Formation and Trading Periods for looping

formation\_length <- 12

trading\_length\_6\_months <- 6

trading\_length\_12\_months <- 12

total\_formation\_trading\_6\_months <- formation\_length + trading\_length\_6\_months

total\_formation\_trading\_12\_months <- formation\_length + trading\_length\_12\_months

formation\_periods <- list()

trading\_periods\_6\_months <- list()

trading\_periods\_12\_months <- list()

years <- index(I1closeprices)[endpoints(I1closeprices, on = "years")]

months <- index(I1closeprices)[endpoints(I1closeprices, on = "months")]

days <- index(I1closeprices)[endpoints(I1closeprices, on = "days")]

npers <- length(years) - 2

for (i in 1:npers){

if(i == 1){

formation\_start <- months[(i-1)\*formation\_length + 1]

}else{

formation\_start <- months[(i-1)\*formation\_length] + days(1)

}

formation\_end <- months[(i-1)\*formation\_length + formation\_length]

formation\_periods[[i]] <- paste(formation\_start, formation\_end, sep = "/")

trading\_start\_6\_months <- formation\_end + days(1)

trading\_end\_6\_months <- trading\_start\_6\_months %m+% months(trading\_length\_6\_months) - days(1)

trading\_periods\_6\_months[[i]] <- paste(trading\_start\_6\_months, trading\_end\_6\_months, sep = "/")

trading\_start\_12\_months <- formation\_end + days(1)

trading\_end\_12\_months <- trading\_start\_12\_months %m+% months(trading\_length\_12\_months) - days(1)

trading\_periods\_12\_months[[i]] <- paste(trading\_start\_12\_months, trading\_end\_12\_months, sep = "/")

}

time\_df <- cbind(formation\_periods, tradingperiod6months = trading\_periods\_6\_months, tradingperiod12months = trading\_periods\_12\_months)

#-------------------------------------------------------------------------------

# Function to process results into interpretable format

process\_results <- function(results\_list, period\_length) {

Returns\_over\_time <- data.frame()

subperiod\_results\_list <- list()

for(i in 1:length(results\_list)){

Period <- names(results\_list)[i]

results\_trading\_period <- results\_list[[Period]][["aggregate\_results\_list"]][[1]]

subperiod\_results\_df <- data.frame(

Method = character(),

Excess\_returns\_on\_actual\_employed\_capital = numeric(),

Monthly\_Excess\_returns\_on\_actual\_employed\_capital = numeric(),

Mean\_return\_per\_trade = numeric(),

Net\_Excess\_returns\_on\_actual\_employed\_capital = numeric(),

Monthly\_Net\_Excess\_returns\_on\_actual\_employed\_capital = numeric(),

Mean\_Net\_returns\_per\_trade = numeric(),

Non\_convergent\_pairs = numeric(),

Proportion\_Non\_Convergent\_Pairs = numeric(),

Average\_Days\_Open\_Per\_Trade = numeric(),

Number\_of\_pairs\_generated = numeric(),

Numer\_of\_pairs\_traded = numeric(),

Loss\_Making\_Pairs = numeric(),

Proportion\_Loss\_Making = numeric(),

stringsAsFactors = FALSE

)

for(j in 1:length(results\_trading\_period)){

element\_name <- names(results\_trading\_period)[j]

if (!startsWith(element\_name, "NULL")){

Excess\_returns\_on\_actual\_employed\_capital <- (sum(results\_trading\_period[[j]][["Total\_Percentage\_Returns"]], na.rm = TRUE) / sum(!is.na(results\_trading\_period[[j]][["Total\_Percentage\_Returns"]])))

Monthly\_Excess\_returns\_on\_actual\_employed\_capital <- Excess\_returns\_on\_actual\_employed\_capital/ period\_length

Non\_convergent\_pairs <- sum(results\_trading\_period[[j]][["Non\_Convergent\_Pair"]], na.rm = TRUE)

Average\_Days\_Open\_Per\_Trade <- mean(results\_trading\_period[[j]][["Average\_Days\_Open\_Per\_Trade"]], na.rm = TRUE)

Number\_of\_pairs\_generated <- nrow(results\_trading\_period[[j]])

Number\_of\_pairs\_traded <- sum(!is.na(results\_trading\_period[[j]][["Total\_Percentage\_Returns"]]))

Proportion\_Non\_Convergent\_Pairs <- Non\_convergent\_pairs/Number\_of\_pairs\_traded

if (any(!is.na(results\_trading\_period[[j]][["Average\_Percentage\_Returns\_Per\_Trade"]]))) {

Mean\_returns\_per\_trade <- mean(results\_trading\_period[[j]][["Average\_Percentage\_Returns\_Per\_Trade"]], na.rm = TRUE)

} else {

Mean\_returns\_per\_trade <- NA

}

Net\_Excess\_returns\_on\_actual\_employed\_capital <- (sum(results\_trading\_period[[j]][["Net\_Total\_Percentage\_Returns"]], na.rm = TRUE) / sum(!is.na(results\_trading\_period[[j]][["Net\_Total\_Percentage\_Returns"]])))

Monthly\_Net\_Excess\_returns\_on\_actual\_employed\_capital <- Net\_Excess\_returns\_on\_actual\_employed\_capital/ period\_length

if (any(!is.na(results\_trading\_period[[j]][["Net\_Average\_Percentage\_Returns\_Per\_Trade"]]))) {

Mean\_Net\_returns\_per\_trade <- mean(results\_trading\_period[[j]][["Net\_Average\_Percentage\_Returns\_Per\_Trade"]], na.rm = TRUE)

} else {

Mean\_Net\_returns\_per\_trade <- NA

}

Loss\_Making\_Pairs <- sum(results\_trading\_period[[j]][["Net\_Total\_Percentage\_Returns"]] < 0, na.rm = TRUE)

Proportion\_Loss\_Making <- Loss\_Making\_Pairs/Number\_of\_pairs\_traded

subperiod\_results\_df <- rbind(subperiod\_results\_df, data.frame(

Method = element\_name,

Excess\_returns\_on\_actual\_employed\_capital = Excess\_returns\_on\_actual\_employed\_capital,

Monthly\_Excess\_returns\_on\_actual\_employed\_capital = Monthly\_Excess\_returns\_on\_actual\_employed\_capital,

Mean\_return\_per\_trade = Mean\_returns\_per\_trade,

Net\_Excess\_returns\_on\_actual\_employed\_capital = Net\_Excess\_returns\_on\_actual\_employed\_capital,

Monthly\_Net\_Excess\_returns\_on\_actual\_employed\_capital = Monthly\_Net\_Excess\_returns\_on\_actual\_employed\_capital,

Mean\_Net\_returns\_per\_trade = Mean\_Net\_returns\_per\_trade,

Non\_convergent\_pairs = Non\_convergent\_pairs,

Proportion\_Non\_Convergent\_Pairs = Proportion\_Non\_Convergent\_Pairs,

Average\_Days\_Open\_Per\_Trade = Average\_Days\_Open\_Per\_Trade,

Number\_of\_pairs\_generated = Number\_of\_pairs\_generated,

Numer\_of\_pairs\_traded = Number\_of\_pairs\_traded,

Loss\_Making\_Pairs = Loss\_Making\_Pairs,

Proportion\_Loss\_Making = Proportion\_Loss\_Making

))

} else {

element\_name <- names(results\_trading\_period)[j]

subperiod\_results\_df <- rbind(subperiod\_results\_df, data.frame(

Method = element\_name,

Excess\_returns\_on\_actual\_employed\_capital = NA,

Monthly\_Excess\_returns\_on\_actual\_employed\_capital = NA,

Mean\_return\_per\_trade = NA,

Net\_Excess\_returns\_on\_actual\_employed\_capital = NA,

Monthly\_Net\_Excess\_returns\_on\_actual\_employed\_capital = NA,

Mean\_Net\_returns\_per\_trade = NA,

Non\_convergent\_pairs = NA,

Proportion\_Non\_Convergent\_Pairs = NA,

Average\_Days\_Open\_Per\_Trade = NA,

Number\_of\_pairs\_generated = NA,

Numer\_of\_pairs\_traded = NA,

Loss\_Making\_Pairs = NA,

Proportion\_Loss\_Making = NA

))

}

}

subperiod\_results\_df$Formation\_Period <- Period

subperiod\_results\_df$Trading\_Period\_Length <- ifelse(period\_length == "6", "6 months", "12 months")

subperiod\_results\_list[[Period]] <- subperiod\_results\_df

}

Returns\_over\_time <- do.call(rbind, subperiod\_results\_list)

rownames(Returns\_over\_time) <- NULL

return(list(Returns\_over\_time = Returns\_over\_time, subperiod\_results\_list = subperiod\_results\_list))

}

#-------------------------------------------------------------------------------

# Script to obtain results, due to lack of processing power this has been seperated into seperate periods

# Standard Deviation Opening thresholds to be determined prior to running the code

Standard\_deviation <- c(2.5, 2.75 , 3, 3.5)

# Script for Generating Results: 6-Month Trading Period Variation, Separated into two year Data Sets over the Observed 6-Year Period

#-------------------

results\_list <- list()

for(t in 1:2){

aggregate\_results\_list <- list()

pair\_information\_list <- list()

formation\_period <- as.character(time\_df[t, "formation\_periods"])

trading\_period\_6 <- as.character(time\_df[t, "tradingperiod6months"])

aggregate\_results\_list\_6 <- list()

pair\_information\_list\_6 <- list()

print(paste("Starting trading period:", trading\_period\_6, "for formation period:", formation\_period))

Distanceformationandtrading <- distancemethod(formation\_period, trading\_period\_6)

Cointegrationfilteredformationandtrading <- cointegrationspeedfilteringmethod(formation\_period, trading\_period\_6)

for (i in seq\_along(Cointegrationfilteredformationandtrading)) {

aspect <- Cointegrationfilteredformationandtrading[[i]]

aspect\_name <- names(Cointegrationfilteredformationandtrading)[i]

if (!is.null(aspect$spread\_analysis\_formation\_filtered\_cointegration)) {

for (sd in Standard\_deviation) {

cointegration\_filtered\_signal\_and\_results <- generate\_signals\_and\_returns(aspect$spread\_analysis\_formation\_filtered\_cointegration, aspect$Cointegrated\_spread\_trading\_filtered\_xts, aspect$normalised\_trading, sd)

filtered\_cointegration\_aggregate\_results <- generate\_aggregate\_results(cointegration\_filtered\_signal\_and\_results$returns\_dataframe\_xts\_cointegration)

aggregate\_results\_list\_6[[paste(aspect\_name, "SD", sd)]] <- filtered\_cointegration\_aggregate\_results

pair\_information\_list\_6[[paste(aspect\_name, "SD", sd)]] <- list(aspect, cointegration\_filtered\_signal\_and\_results, filtered\_cointegration\_aggregate\_results)

}

} else {

for (sd in Standard\_deviation) {

message <- paste("No pairs above the speed threshold for", aspect\_name, "SD", sd)

aggregate\_results\_list\_6[[paste("NULL", aspect\_name, "SD", sd)]] <- list(message)

pair\_information\_list\_6[[paste("NULL", aspect\_name, "SD", sd)]] <- list(NULL, NULL, NULL)

print(message)

}

}

}

for(sd in Standard\_deviation) {

distance\_signal\_and\_results <- generate\_signals\_and\_returns(Distanceformationandtrading$Spread\_Analysis\_Formation\_Distance, Distanceformationandtrading$Distance\_spread\_trading\_xts, Distanceformationandtrading$normalised\_trading, sd )

distance\_aggregate\_results <- generate\_aggregate\_results(distance\_signal\_and\_results$returns\_dataframe\_xts\_Distance)

aggregate\_results\_list\_6[[paste("Distance Method SD", sd)]] <- distance\_aggregate\_results

pair\_information\_list\_6[[paste("Distance Method SD", sd)]] <- list(Distanceformationandtrading, distance\_signal\_and\_results, distance\_aggregate\_results)

}

print(" 6 Month Trading Period Complete")

aggregate\_results\_list[[as.character(trading\_period\_6)]] <- aggregate\_results\_list\_6

pair\_information\_list[[as.character(trading\_period\_6)]] <- pair\_information\_list\_6

results\_list[[as.character(formation\_period)]] <- list(aggregate\_results\_list = aggregate\_results\_list, pair\_information\_list = pair\_information\_list)

print("Formation Loop Complete")

}

gc()

# Gc() is necessary to prevent the code freezing

save(results\_list, file = "C:/Dissertation/6monthsresults3and4period.RData")

# Results are processed straight away to save data

AG612 <- process\_results(results\_list, 6)

save(AG612, file = "C:/Dissertation/AG612")

results\_list <- list()

#-------------------

for(t in 3:4){

aggregate\_results\_list <- list()

pair\_information\_list <- list()

formation\_period <- as.character(time\_df[t, "formation\_periods"])

trading\_period\_6 <- as.character(time\_df[t, "tradingperiod6months"])

aggregate\_results\_list\_6 <- list()

pair\_information\_list\_6 <- list()

print(paste("Starting trading period:", trading\_period\_6, "for formation period:", formation\_period))

Distanceformationandtrading <- distancemethod(formation\_period, trading\_period\_6)

Cointegrationfilteredformationandtrading <- cointegrationspeedfilteringmethod(formation\_period, trading\_period\_6)

for (i in seq\_along(Cointegrationfilteredformationandtrading)) {

aspect <- Cointegrationfilteredformationandtrading[[i]]

aspect\_name <- names(Cointegrationfilteredformationandtrading)[i]

if (!is.null(aspect$spread\_analysis\_formation\_filtered\_cointegration)) {

for (sd in Standard\_deviation) {

cointegration\_filtered\_signal\_and\_results <- generate\_signals\_and\_returns(aspect$spread\_analysis\_formation\_filtered\_cointegration, aspect$Cointegrated\_spread\_trading\_filtered\_xts, aspect$normalised\_trading, sd)

filtered\_cointegration\_aggregate\_results <- generate\_aggregate\_results(cointegration\_filtered\_signal\_and\_results$returns\_dataframe\_xts\_cointegration)

aggregate\_results\_list\_6[[paste(aspect\_name, "SD", sd)]] <- filtered\_cointegration\_aggregate\_results

pair\_information\_list\_6[[paste(aspect\_name, "SD", sd)]] <- list(aspect, cointegration\_filtered\_signal\_and\_results, filtered\_cointegration\_aggregate\_results)

}

} else {

for (sd in Standard\_deviation) {

message <- paste("No pairs above the speed threshold for", aspect\_name, "SD", sd)

aggregate\_results\_list\_6[[paste("NULL", aspect\_name, "SD", sd)]] <- list(message)

pair\_information\_list\_6[[paste("NULL", aspect\_name, "SD", sd)]] <- list(NULL, NULL, NULL)

print(message)

}

}

}

for(sd in Standard\_deviation) {

distance\_signal\_and\_results <- generate\_signals\_and\_returns(Distanceformationandtrading$Spread\_Analysis\_Formation\_Distance, Distanceformationandtrading$Distance\_spread\_trading\_xts, Distanceformationandtrading$normalised\_trading, sd )

distance\_aggregate\_results <- generate\_aggregate\_results(distance\_signal\_and\_results$returns\_dataframe\_xts\_Distance)

aggregate\_results\_list\_6[[paste("Distance Method SD", sd)]] <- distance\_aggregate\_results

pair\_information\_list\_6[[paste("Distance Method SD", sd)]] <- list(Distanceformationandtrading, distance\_signal\_and\_results, distance\_aggregate\_results)

}

print(" 6 Month Trading Period Complete")

aggregate\_results\_list[[as.character(trading\_period\_6)]] <- aggregate\_results\_list\_6

pair\_information\_list[[as.character(trading\_period\_6)]] <- pair\_information\_list\_6

results\_list[[as.character(formation\_period)]] <- list(aggregate\_results\_list = aggregate\_results\_list, pair\_information\_list = pair\_information\_list)

print("Formation Loop Complete")

}

gc()

save(results\_list, file = "C:/Dissertation/6monthsresults3and4period.RData")

AG634 <- process\_results(results\_list, 6)

save(AG634, file = "C:/Dissertation/AG634")

results\_list <- list()

#-------------------

for(t in 5:6){

aggregate\_results\_list <- list()

pair\_information\_list <- list()

formation\_period <- as.character(time\_df[t, "formation\_periods"])

trading\_period\_6 <- as.character(time\_df[t, "tradingperiod6months"])

aggregate\_results\_list\_6 <- list()

pair\_information\_list\_6 <- list()

print(paste("Starting trading period:", trading\_period\_6, "for formation period:", formation\_period))

Distanceformationandtrading <- distancemethod(formation\_period, trading\_period\_6)

Cointegrationfilteredformationandtrading <- cointegrationspeedfilteringmethod(formation\_period, trading\_period\_6)

for (i in seq\_along(Cointegrationfilteredformationandtrading)) {

aspect <- Cointegrationfilteredformationandtrading[[i]]

aspect\_name <- names(Cointegrationfilteredformationandtrading)[i]

if (!is.null(aspect$spread\_analysis\_formation\_filtered\_cointegration)) {

for (sd in Standard\_deviation) {

cointegration\_filtered\_signal\_and\_results <- generate\_signals\_and\_returns(aspect$spread\_analysis\_formation\_filtered\_cointegration, aspect$Cointegrated\_spread\_trading\_filtered\_xts, aspect$normalised\_trading, sd)

filtered\_cointegration\_aggregate\_results <- generate\_aggregate\_results(cointegration\_filtered\_signal\_and\_results$returns\_dataframe\_xts\_cointegration)

aggregate\_results\_list\_6[[paste(aspect\_name, "SD", sd)]] <- filtered\_cointegration\_aggregate\_results

pair\_information\_list\_6[[paste(aspect\_name, "SD", sd)]] <- list(aspect, cointegration\_filtered\_signal\_and\_results, filtered\_cointegration\_aggregate\_results)

}

} else {

for (sd in Standard\_deviation) {

message <- paste("No pairs above the speed threshold for", aspect\_name, "SD", sd)

aggregate\_results\_list\_6[[paste("NULL", aspect\_name, "SD", sd)]] <- list(message)

pair\_information\_list\_6[[paste("NULL", aspect\_name, "SD", sd)]] <- list(NULL, NULL, NULL)

print(message)

}

}

}

for(sd in Standard\_deviation) {

distance\_signal\_and\_results <- generate\_signals\_and\_returns(Distanceformationandtrading$Spread\_Analysis\_Formation\_Distance, Distanceformationandtrading$Distance\_spread\_trading\_xts, Distanceformationandtrading$normalised\_trading, sd )

distance\_aggregate\_results <- generate\_aggregate\_results(distance\_signal\_and\_results$returns\_dataframe\_xts\_Distance)

aggregate\_results\_list\_6[[paste("Distance Method SD", sd)]] <- distance\_aggregate\_results

pair\_information\_list\_6[[paste("Distance Method SD", sd)]] <- list(Distanceformationandtrading, distance\_signal\_and\_results, distance\_aggregate\_results)

}

print(" 6 Month Trading Period Complete")

aggregate\_results\_list[[as.character(trading\_period\_6)]] <- aggregate\_results\_list\_6

pair\_information\_list[[as.character(trading\_period\_6)]] <- pair\_information\_list\_6

results\_list[[as.character(formation\_period)]] <- list(aggregate\_results\_list = aggregate\_results\_list, pair\_information\_list = pair\_information\_list)

print("Formation Loop Complete")

}

gc()

save(results\_list, file = "C:/Dissertation/6monthsresultsperiods5and6.RData")

AG656 <- process\_results(results\_list, 6)

save(AG634, file = "C:/Dissertation/AG656")

#-------------------------------------------------------------------------------

# Script for Generating Results: 12-Month Trading Period Variation, Separated into two Data Sets over the Observed 6-Year Period

results\_list <- list()

for(t in 1:2){

aggregate\_results\_list <- list()

pair\_information\_list <- list()

formation\_period <- as.character(time\_df[t, "formation\_periods"])

trading\_period\_12 <- as.character(time\_df[t, "tradingperiod12months"])

aggregate\_results\_list\_12 <- list()

pair\_information\_list\_12 <- list()

print(paste("Starting trading period:", trading\_period\_12, "for formation period:", formation\_period))

Distanceformationandtrading <- distancemethod(formation\_period, trading\_period\_12)

Cointegrationfilteredformationandtrading <- cointegrationspeedfilteringmethod(formation\_period, trading\_period\_12)

for (i in seq\_along(Cointegrationfilteredformationandtrading)) {

aspect <- Cointegrationfilteredformationandtrading[[i]]

aspect\_name <- names(Cointegrationfilteredformationandtrading)[i]

if (!is.null(aspect$spread\_analysis\_formation\_filtered\_cointegration)) {

for (sd in Standard\_deviation) {

cointegration\_filtered\_signal\_and\_results <- generate\_signals\_and\_returns(aspect$spread\_analysis\_formation\_filtered\_cointegration, aspect$Cointegrated\_spread\_trading\_filtered\_xts, aspect$normalised\_trading, sd)

filtered\_cointegration\_aggregate\_results <- generate\_aggregate\_results(cointegration\_filtered\_signal\_and\_results$returns\_dataframe\_xts\_cointegration)

aggregate\_results\_list\_12[[paste(aspect\_name, "SD", sd)]] <- filtered\_cointegration\_aggregate\_results

pair\_information\_list\_12[[paste(aspect\_name, "SD", sd)]] <- list(aspect, cointegration\_filtered\_signal\_and\_results, filtered\_cointegration\_aggregate\_results)

}

} else {

for (sd in Standard\_deviation) {

message <- paste("No pairs above the speed threshold for", aspect\_name, "SD", sd)

aggregate\_results\_list\_12[[paste("NULL", aspect\_name, "SD", sd)]] <- list(message)

pair\_information\_list\_12[[paste("NULL", aspect\_name, "SD", sd)]] <- list(NULL, NULL, NULL)

print(message)

}

}

}

for(sd in Standard\_deviation) {

distance\_signal\_and\_results <- generate\_signals\_and\_returns(Distanceformationandtrading$Spread\_Analysis\_Formation\_Distance, Distanceformationandtrading$Distance\_spread\_trading\_xts, Distanceformationandtrading$normalised\_trading, sd )

distance\_aggregate\_results <- generate\_aggregate\_results(distance\_signal\_and\_results$returns\_dataframe\_xts\_Distance)

aggregate\_results\_list\_12[[paste("Distance Method SD", sd)]] <- distance\_aggregate\_results

pair\_information\_list\_12[[paste("Distance Method SD", sd)]] <- list(Distanceformationandtrading, distance\_signal\_and\_results, distance\_aggregate\_results)

}

print(" 6 Month Trading Period Complete")

aggregate\_results\_list[[as.character(trading\_period\_12)]] <- aggregate\_results\_list\_12

pair\_information\_list[[as.character(trading\_period\_12)]] <- pair\_information\_list\_12

results\_list[[as.character(formation\_period)]] <- list(aggregate\_results\_list = aggregate\_results\_list, pair\_information\_list = pair\_information\_list)

print("Formation Loop Complete")

}

gc()

save(results\_list, file = "C:/Dissertation/12monthsresultsfirst2p.RData")

AG1212 <- process\_results(results\_list, 12)

save(AG1212, file = "C:/Dissertation/AG1212")

results\_list <- list()

#-------------------

for(t in 3:4){

aggregate\_results\_list <- list()

pair\_information\_list <- list()

formation\_period <- as.character(time\_df[t, "formation\_periods"])

trading\_period\_12 <- as.character(time\_df[t, "tradingperiod12months"])

aggregate\_results\_list\_12 <- list()

pair\_information\_list\_12 <- list()

print(paste("Starting trading period:", trading\_period\_12, "for formation period:", formation\_period))

Distanceformationandtrading <- distancemethod(formation\_period, trading\_period\_12)

Cointegrationfilteredformationandtrading <- cointegrationspeedfilteringmethod(formation\_period, trading\_period\_12)

for (i in seq\_along(Cointegrationfilteredformationandtrading)) {

aspect <- Cointegrationfilteredformationandtrading[[i]]

aspect\_name <- names(Cointegrationfilteredformationandtrading)[i]

if (!is.null(aspect$spread\_analysis\_formation\_filtered\_cointegration)) {

for (sd in Standard\_deviation) {

cointegration\_filtered\_signal\_and\_results <- generate\_signals\_and\_returns(aspect$spread\_analysis\_formation\_filtered\_cointegration, aspect$Cointegrated\_spread\_trading\_filtered\_xts, aspect$normalised\_trading, sd)

filtered\_cointegration\_aggregate\_results <- generate\_aggregate\_results(cointegration\_filtered\_signal\_and\_results$returns\_dataframe\_xts\_cointegration)

aggregate\_results\_list\_12[[paste(aspect\_name, "SD", sd)]] <- filtered\_cointegration\_aggregate\_results

pair\_information\_list\_12[[paste(aspect\_name, "SD", sd)]] <- list(aspect, cointegration\_filtered\_signal\_and\_results, filtered\_cointegration\_aggregate\_results)

}

} else {

for (sd in Standard\_deviation) {

message <- paste("No pairs above the speed threshold for", aspect\_name, "SD", sd)

aggregate\_results\_list\_12[[paste("NULL", aspect\_name, "SD", sd)]] <- list(message)

pair\_information\_list\_12[[paste("NULL", aspect\_name, "SD", sd)]] <- list(NULL, NULL, NULL)

print(message)

}

}

}

for(sd in Standard\_deviation) {

distance\_signal\_and\_results <- generate\_signals\_and\_returns(Distanceformationandtrading$Spread\_Analysis\_Formation\_Distance, Distanceformationandtrading$Distance\_spread\_trading\_xts, Distanceformationandtrading$normalised\_trading, sd )

distance\_aggregate\_results <- generate\_aggregate\_results(distance\_signal\_and\_results$returns\_dataframe\_xts\_Distance)

aggregate\_results\_list\_12[[paste("Distance Method SD", sd)]] <- distance\_aggregate\_results

pair\_information\_list\_12[[paste("Distance Method SD", sd)]] <- list(Distanceformationandtrading, distance\_signal\_and\_results, distance\_aggregate\_results)

}

print(" 6 Month Trading Period Complete")

aggregate\_results\_list[[as.character(trading\_period\_12)]] <- aggregate\_results\_list\_12

pair\_information\_list[[as.character(trading\_period\_12)]] <- pair\_information\_list\_12

results\_list[[as.character(formation\_period)]] <- list(aggregate\_results\_list = aggregate\_results\_list, pair\_information\_list = pair\_information\_list)

print("Formation Loop Complete")

}

gc()

save(results\_list, file = "C:/Dissertation/12monthsresults3and4period.RData")

AG1234 <- process\_results(results\_list, 12)

save(AG1234, file = "C:/Dissertation/AG1234")

results\_list <- list()

#-------------------

for(t in 5:6){

aggregate\_results\_list <- list()

pair\_information\_list <- list()

formation\_period <- as.character(time\_df[t, "formation\_periods"])

trading\_period\_12 <- as.character(time\_df[t, "tradingperiod12months"])

aggregate\_results\_list\_12 <- list()

pair\_information\_list\_12 <- list()

print(paste("Starting trading period:", trading\_period\_12, "for formation period:", formation\_period))

Distanceformationandtrading <- distancemethod(formation\_period, trading\_period\_12)

Cointegrationfilteredformationandtrading <- cointegrationspeedfilteringmethod(formation\_period, trading\_period\_12)

for (i in seq\_along(Cointegrationfilteredformationandtrading)) {

aspect <- Cointegrationfilteredformationandtrading[[i]]

aspect\_name <- names(Cointegrationfilteredformationandtrading)[i]

if (!is.null(aspect$spread\_analysis\_formation\_filtered\_cointegration)) {

for (sd in Standard\_deviation) {

cointegration\_filtered\_signal\_and\_results <- generate\_signals\_and\_returns(aspect$spread\_analysis\_formation\_filtered\_cointegration, aspect$Cointegrated\_spread\_trading\_filtered\_xts, aspect$normalised\_trading, sd)

filtered\_cointegration\_aggregate\_results <- generate\_aggregate\_results(cointegration\_filtered\_signal\_and\_results$returns\_dataframe\_xts\_cointegration)

aggregate\_results\_list\_12[[paste(aspect\_name, "SD", sd)]] <- filtered\_cointegration\_aggregate\_results

pair\_information\_list\_12[[paste(aspect\_name, "SD", sd)]] <- list(aspect, cointegration\_filtered\_signal\_and\_results, filtered\_cointegration\_aggregate\_results)

}

} else {

for (sd in Standard\_deviation) {

message <- paste("No pairs above the speed threshold for", aspect\_name, "SD", sd)

aggregate\_results\_list\_12[[paste("NULL", aspect\_name, "SD", sd)]] <- list(message)

pair\_information\_list\_12[[paste("NULL", aspect\_name, "SD", sd)]] <- list(NULL, NULL, NULL)

print(message)

}

}

}

for(sd in Standard\_deviation) {

distance\_signal\_and\_results <- generate\_signals\_and\_returns(Distanceformationandtrading$Spread\_Analysis\_Formation\_Distance, Distanceformationandtrading$Distance\_spread\_trading\_xts, Distanceformationandtrading$normalised\_trading, sd )

distance\_aggregate\_results <- generate\_aggregate\_results(distance\_signal\_and\_results$returns\_dataframe\_xts\_Distance)

aggregate\_results\_list\_12[[paste("Distance Method SD", sd)]] <- distance\_aggregate\_results

pair\_information\_list\_12[[paste("Distance Method SD", sd)]] <- list(Distanceformationandtrading, distance\_signal\_and\_results, distance\_aggregate\_results)

}

print(" 6 Month Trading Period Complete")

aggregate\_results\_list[[as.character(trading\_period\_12)]] <- aggregate\_results\_list\_12

pair\_information\_list[[as.character(trading\_period\_12)]] <- pair\_information\_list\_12

results\_list[[as.character(formation\_period)]] <- list(aggregate\_results\_list = aggregate\_results\_list, pair\_information\_list = pair\_information\_list)

print("Formation Loop Complete")

}

gc()

save(results\_list, file = "C:/Dissertation/12monthsresultsperiods5and6.RData")

AG1256 <- process\_results(results\_list, 12)

save(AG1256, file = "C:/Dissertation/AG1256")

#-------------------------------------------------------------------------------

# Subperiod Results are merged to calculate means over time and arranged into interpretable format

Aggregate\_Results\_Combined <- do.call(rbind, lapply(list(AG612,

AG634,

AG656,

AG1212,

AG1234,

AG1256),

function(x) x$Returns\_over\_time))

rownames(Aggregate\_Results\_Combined) <- NULL

write.csv(Aggregate\_Results\_Combined, file = "C:/Final Aggregate/PostconclusionAggregate")

finalresults <- Aggregate\_Results\_Combined %>%

filter(!startsWith(Method, "NULL")) %>%

separate(Method, into = c("Method", "SD"), sep = " SD ", remove = TRUE) %>%

group\_by(Method,SD, Trading\_Period\_Length) %>%

summarize\_all(mean) %>%

select(-Formation\_Period)

refineddata <- finalresults %>%

select(Trading\_Period\_Length, Method, SD, contains("Monthly"), Proportion\_Non\_Convergent\_Pairs, Proportion\_Loss\_Making, Number\_of\_pairs\_generated)

#write.csv(refindeddata file = "")

#-------------------------------------------------------------------------------

# Script used to run the linear regression referenced in the Dissertation

library(lmtest)

library(car)

library(ggplot2)

library(dplyr)

library(broom)

library(ggpubr)

regression\_subset <- stoploss1.5mean[c("Proportion\_Loss\_Making\_mean", "Proportion\_Non\_Convergent\_Pairs\_mean")]

regression <- lm(Proportion\_Loss\_Making\_mean ~ Proportion\_Non\_Convergent\_Pairs\_mean, data = regression\_subset)

summary(regression)

bptest(regression)

residualPlot(regression)

#-------------------------------------------------------------------------------

# Function to obtain Graphs

graphofstocksandsignals <- function(Aggregate\_results, formationandtradingdata, signaldata){

list\_of\_plots\_and\_trade\_positions=list()

for (i in 1:nrow(Aggregate\_results)){

if (anyNA(Aggregate\_results[i, ])) {

next

}

pairjointname <- Aggregate\_results[i,1]

pair <- unlist(strsplit(pairjointname, "\_"))

stock1 <- as.character(pair[1])

stock2 <- as.character(pair[2])

plot <- chart\_Series(formationandtradingdata[["I1closeprices\_trading"]][, stock1], name = stock1, col = "red")

add\_TA(formationandtradingdata[["I1closeprices\_trading"]][, stock2], on = 1, col = "blue", lty = "dashed")

add\_TA(signaldata[[1]][[i]][,"nsignal"], type = 's', col = 'red')

chart$ylim <- c(-1.5, 1.5)

list\_of\_plots\_and\_trade\_positions[[i]] <- plot

}

return(list\_of\_plots\_and\_trade\_positions)

}

#Input Format: Inputs should be provided in long form from the results\_list datafram.

cointegration\_graphs\_of\_stock\_and\_signals <- graphofstocksandsignals(Cointegration\_Aggregate\_Results,Cointegrationformationandtrading, cointegration\_signal\_and\_results)

#-------------------------------------------------------------------------------

# Function to obtain a graph of the Spreads

graphof\_spread\_signal <- function(Aggregate\_results, signaldata){

list\_of\_spreads\_and\_trade\_positions=list()

for (i in 1:nrow(Aggregate\_results)){

if (anyNA(Aggregate\_results[i, ])) {

next

}

pairspreadname <- paste(Aggregate\_results[i, 1], "\_spread", sep = "")

plot <- chart\_Series(signaldata[[1]][[i]][,pairspreadname], col = 'blue')

add\_TA(signaldata[[1]][[i]][,"nsignal"], type = 's', col = 'red')

list\_of\_spreads\_and\_trade\_positions[[i]] <- plot

}

return(list\_of\_spreads\_and\_trade\_positions)

}

# Input Format: Inputs should be provided in long form from the results\_list dataframe

cointegrationspreadsignal <- graphof\_spread\_signal(Cointegration\_Aggregate\_Results, cointegration\_signal\_and\_results)

#-------------------------------------------------------------------------------

REFERENCE LIST

Caldeira, J. and Moura, G.V. (2013). Selection of a Portfolio of Pairs Based on Cointegration: A Statistical Arbitrage Strategy. [online] papers.ssrn.com. Available at: https://papers.ssrn.com/sol3/papers.cfm?abstract\_id=2196391.

Carrasco Blázquez, M., De la Orden De la Cruz, C. and Prado Román, C. (2018). Pairs trading techniques: An empirical contrast. European Research on Management and Business Economics, 24(3), pp.160–167. doi:https://doi.org/10.1016/j.iedeen.2018.05.002.

Do, B. and Faff, R. (2010). Does Simple Pairs Trading Still Work? Financial Analysts Journal, 66(4), pp.83–95. doi:https://doi.org/10.2469/faj.v66.n4.1. ddo

Do, B. and Faff, R. (2012). ARE PAIRS TRADING PROFITS ROBUST TO TRADING COSTS? Journal of Financial Research, 35(2), pp.261–287. doi:https://doi.org/10.1111/j.1475-6803.2012.01317.x.

Emerson, J. (2007). Cointegration analysis and the choice of lag length. Applied Economics Letters, [online] 14(12), pp.881–885. Available at: https://ideas.repec.org/a/taf/apeclt/v14y2007i12p881-885.html [Accessed 18 Apr. 2024].

Emery, G.W. and Liu, Q. (Wilson) (2002). An analysis of the relationship between electricity and natural-gas futures prices. Journal of Futures Markets, 22(2), p.95. doi:https://doi.org/10.1002/fut.2209.

Enders, W. (2014). Applied Econometric Times Series. Wiley.

Engle, R.F. and Granger, C.W.J. (1987). Co-Integration and Error Correction: Representation, Estimation, and Testing. Econometrica, 55(2), pp.251–276. doi:https://doi.org/10.2307/1913236.

Gatev, E., Goetzmann, W. and Rouwenhorst, K.G. (1999). Pairs Trading: Performance of a Relative Value Arbitrage Rule. doi:https://doi.org/10.3386/w7032.

Gatev, E., Goetzmann, W.N. and Rouwenhorst, K.G. (2006). Pairs Trading: Performance of a Relative Value Arbitrage Rule. [online] papers.ssrn.com. Available at: https://papers.ssrn.com/sol3/papers.cfm?abstract\_id=141615.

Ganapathy Vidyamurthy (2011). Pairs Trading. John Wiley & Sons.

Girma, P.B. and Paulson, A.S. (1999). Risk arbitrage opportunities in petroleum futures spreads.

Interactive Brokers (2024). Commissions Stocks | Interactive Brokers U.K. Limited. [online] www.interactivebrokers.co.uk. Available at: https://www.interactivebrokers.co.uk/en/pricing/commissions-stocks.php [Accessed 18 Apr. 2024].

Interactive Brokers (2024b). Short Sale Cost | Interactive Brokers U.K. Limited. [online] www.interactivebrokers.co.uk. Available at: https://www.interactivebrokers.co.uk/en/pricing/short-sale-cost.php [Accessed 18 Apr. 2024].

Journal of Futures Markets, [online] 19(8), pp.931–955. Available at: https://econpapers.repec.org/article/wlyjfutmk/v\_3a19\_3ay\_3a1999\_3ai\_3a8\_3ap\_3a931-955.htm [Accessed 19 Dec. 2023].

Granger, C.W.J. and Newbold, P. (1974). Spurious regressions in econometrics. Journal of Econometrics, [online] 2(2), pp.111–120. doi:https://doi.org/10.1016/0304-4076(74)90034-7.

Granger, C.W.J. and Weiss, A.A. (1983). TIME SERIES ANALYSIS OF ERROR-CORRECTION MODELS. [online] ScienceDirect. Available at: https://www.sciencedirect.com/science/article/abs/pii/B9780123987501500188 [Accessed 19 Dec. 2023].

Gulati, C., Lin, Y.-X. and McCrae, M. (2006). Loss protection in pairs trading through minimum profit bounds: a Cointegration approach. Faculty of Informatics - Papers (Archive), [online] p.Art. No. 73803. doi:https://doi.org/10.1155/JAMDS/2006/73803.

Huck, N. (2013). The high sensitivity of pairs trading returns. Applied Economics Letters, 20(14), pp.1301–1304. doi:https://doi.org/10.1080/13504851.2013.802121.

Huck, N. and Afawubo, K. (2014). Pairs trading and selection methods: is Cointegration superior? Applied Economics, 47(6), pp.599–613. doi:https://doi.org/10.1080/00036846.2014.975417.

Johansen, S. (1988). Statistical analysis of Cointegration vectors. Journal of Economic Dynamics and Control, 12(2-3), pp.231–254.

Miao, G.J. (2014). High Frequency and Dynamic Pairs Trading Based on Statistical Arbitrage Using a Two-Stage Correlation and Cointegration Approach. International Journal of Economics and Finance, 6(3). doi:https://doi.org/10.5539/ijef.v6n3p96.

Nair, S.T.G. (2021). Price discovery and pairs trading potentials: the case of metals markets. Journal of Financial Economic Policy, [online] 13(5), pp.565–586. Available at: https://ideas.repec.org/a/eme/jfeppp/jfep-06-2020-0139.html [Accessed 19 Dec. 2023].

Nasdaq (2021). High Frequency Trading (HFT) Definition. [online] Nasdaq.com. Available at: https://www.nasdaq.com/glossary/h/high-frequency-trading#:~:text=It%20is%20estimated%20that%2050 [Accessed 18 Apr. 2024].

Rad, H., Low, R.K.Y. and Faff, R.W. (2015). The Profitability of Pairs Trading Strategies: Distance, Cointegration, and Copula Methods. SSRN Electronic Journal. [online] doi:https://doi.org/10.2139/ssrn.2614233.

Simon, D.P. (1999). The soybean crush spread: Empirical evidence and trading strategies. Journal of Futures Markets, 19(3), pp.271–289. doi:https://doi.org/10.1002/(sici)1096-9934(199905)19:3%3C271::aid-fut2%3E3.0.co;2-p.

Taskeen Qazi, L., Rahman, A. and Gul, S. (2015). Which Pairs of Stocks should we Trade? Selection of Pairs for Statistical Arbitrage and Pairs Trading in Karachi Stock Exchange. The Pakistan Development Review, [online] 54(3), pp.215–244. Available at: https://econpapers.repec.org/article/pidjournl/v\_3a54\_3ay\_3a2015\_3ai\_3a3\_3ap\_3a215-244.htm [Accessed 19 Dec. 2023].