

Statistical Arbitrage: High Frequency Pairs Trading

Ruben Joakim Gundersen

Supervisor: Associate Professor Michael Kissler

NORWEGIAN SCHOOL OF ECONOMICS

Master Thesis, MSc. Economics and Business Administration

Financial Economics

This thesis was written as a part of the Master of Science in Economics and Business Administration at NHH. Please note that neither the institution nor the examiners are responsible – through the approval of this thesis – for the theories and methods used, or results and conclusions drawn in this work.

Abstract

In this thesis we examine the performance of a relative value strategy called Pairs Trading. Pairs Trading is one of several strategies collectively referred to as *Statistical Arbitrage strategies*. Candidate pairs are formed by matching stocks with similar historical price paths. The pairs, once matched, are automatically traded based on a set of trading rules. We conduct an empirical analysis using high frequency intraday data from the first quarter of 2014. Our findings indicate that the strategy is able to generate positive risk adjusted returns, even after controlling for moderate transaction costs and placing constraints on the speed of order execution.

Preface

This thesis marks the end of my master studies at NHH. The work has at times been challenging, but at the same time it has also been very rewarding. I would like to thank my supervisor, Michael Kisser, for valuable input and suggestions during the writing process.

Contents

1. Introduction	1
2. The concept of pairs trading	3
2.1 Normalization of stock prices	5
3. Pairs trading in previous literature.....	7
4. Different approaches to pairs trading.....	11
4.1 The Cointegration approach	11
4.2 The Distance approach	15
4.3 The Stochastic approach	16
5. Simulation testing and choice of approach	19
5.1 Model one - The granger representation theorem.....	20
5.1.1 Results.....	22
5.2 Model two - The Stock & Watson Common trends model.....	23
5.2.1 Results.....	27
5.3 Consequences of inaccurate cointegration coefficient estimates	27
5.4 Choice of method.....	28
6 Testing a high frequency pairs trading strategy.....	29
6.1 Data.....	29
6.2 Formation and trading periods.....	31
6.3 Practical Implementation of the distance approach	32
6.3.1 Pairs formation	32
6.3.2 Pairs Trading	32
6.4 Return computation	34
6.5 An algorithmic representation of the test setup	35
6.6 Results.....	36
6.6.1 Unrestricted pairs matching	37
6.6.2 Restricted case – Pairs from corresponding industries.....	39
6.7 Impact of transaction costs and timing constraints.....	42
6.7.1 Commissions and short fees	43
6.7.2 Speed of execution.....	45
6.7.3 Concluding remarks commissions and execution speed	47
6.6.4 Trade slippage.....	48

6.8	Exposure to market risk	49
6.9	Is pairs trading a masked mean reversion strategy?	51
7.	Conclusion.....	53
	References.....	53
	Appendices.....	55

1. Introduction

In this paper we examine a popular quantitative investment strategy commonly referred to as “pairs trading”. The basic concept of pairs trading is remarkably simple; one identifies a pair of stocks that exhibit historical co-movement in prices. Subsequently, if significant deviations from the historical relationship are observed, a position is opened. The position is formed by simultaneously selling short the relative winner and buying long the relative loser. When the prices eventually converge the position is closed and a profit is made. The strategy builds upon the notion that the relative prices in a market are in equilibrium, and that deviations from this equilibrium eventually will be corrected. Applying a pairs trading strategy is therefore an attempt to profit from temporary deviations from this equilibrium.

According to Gatev, Goetzmann & Rouwenhorst (2006) pairs trading strategies have been used by practitioners on Wall Street in various forms since the mid-1980s. The strategy is often said to have originated within Morgan Stanley in a group led by Nunzio Tartaglia. The focus of the group was to develop quantitative trading strategies by employing advanced statistical models and information technology. The group sought to “mechanize” the investment process by developing trading rules that could be automated. Pairs trading was one of the resulting strategies. The group used this strategy with great success in 1987 – when the group is said to have generated a profit of \$50 million – but were dissolved in 1989 after a period of poor performance. In the last decades, as technology has become more accessible, the strategy has been increasingly popular with investors.

Pairs trading is often placed in a group of quantitative trading approaches collectively referred to as statistical arbitrage strategies. The arbitrage part in this context is somewhat misleading as arbitrage implies a risk free profit opportunity at zero upfront cost. A pairs trading strategy is by no means risk free. There is no guarantee that the stocks in a pair will converge. They could even continue to

diverge, resulting in significant losses. Furthermore, the strategy is also often claimed to be “market-neutral”, meaning that the investor is unexposed to the general market risk. However, while it certainly is possible to create market neutral pairs, the total market risk of a position depends on the amount of capital placed in each stock and the sensitivity of the stocks to such risk.

*

In the first part of this thesis we explore the background and the theoretical basis for a pairs trading strategy. In addition we compare the performance of two existing pairs trading methods by applying them to sets of simulated data.

In the latter part of the paper we conduct an empirical analysis of a concrete pairs trading strategy. Through this analysis, we seek to determine if a pairs trading strategy delivers returns that are superior when compared to a buy-and-hold strategy. We use high frequency data for stocks listed on the Oslo Stock Exchange. The obtained results indicate that it is possible to generate positive risk-adjusted returns by following a pairs trading strategy. The results are robust after controlling for transaction costs, and placing restrictions on the execution speed. Specifically, we report annualized returns as high as 12 after costs %. In addition, the standard deviations of the returns are low. This combination leads to an impressive Sharpe ratio exceeding 3. We find that the constructed portfolios have close to zero exposure to market risk.

2. The concept of pairs trading

The pairs trading strategy is based on the concept of relative pricing. If two securities have identical payoffs in all states their price should also be identical. This is a variant of the principle commonly referred to as the Law of One price (LOP). Lamont and Thaler (2003, 191) defines the LOP as follows “[...] *identical goods must have identical prices*”. It is important to note that the prices do not need to be “correct”, from an economical point of view, for the LOP to be valid. The LOP simply asserts that stocks yielding identical payoffs should have the same current price. The law is therefore applicable to the relative pricing of the stocks in a market, even if the pricing is economically incorrect (Gatev et al., 2008). We can further extend the example with identical payoffs to a situation where the payoffs are very similar but not identical. In such a situation the prices of the securities should also be similar. If a temporary deviation from this relative pricing relationship occurs it should be possible to exploit this by taking a position that generates a profit when the deviation is corrected. Pairs trading is one example of a strategy aiming to profit from such temporary deviations.

Before a pairs trading strategy can be implemented on a practical level we need to address some fundamental questions: What pairs of stocks are suitable? When should a position be opened or closed? How should one determine the amount of capital placed in the individual long/short positions? As we will see in section 4, there are multiple approaches to pairs trading, all offering different answers to these questions. Even so, the basic structure of a pairs trading strategy is common for all approaches. The first step involves identifying a pair of stocks whose prices appear to move together according to some fixed relationship. The period of time used to establish such a relationship is referred to as the formation period. After the suitable pairs are identified we enter the trading period. In this period we continue to observe the spread. If a significant deviation from the relationship is observed a position is opened. The investor then buys long some quantity of the relative loser and sells short some quantity of the relative winner.

The following figure graphically illustrates the concept of the pairs trading strategy.

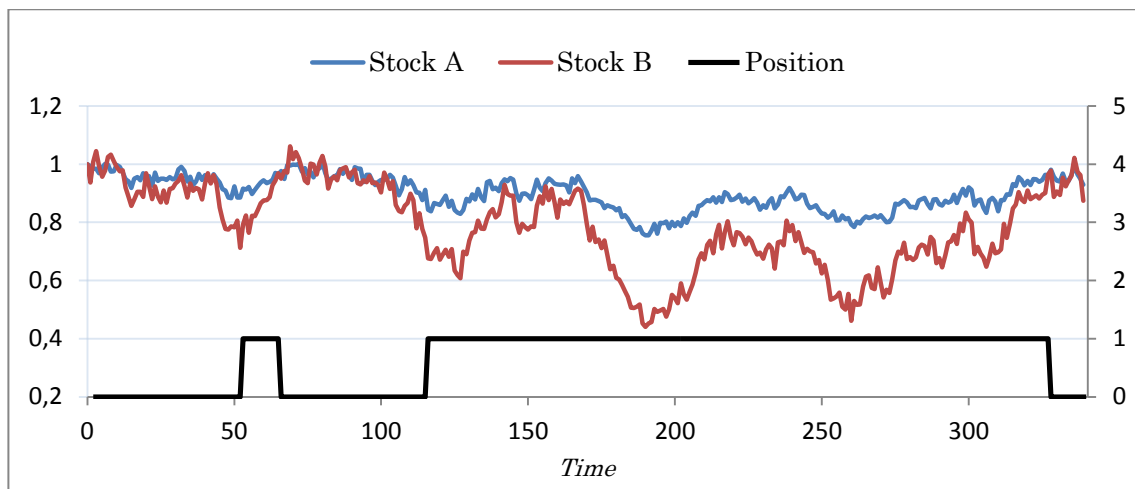


Figure 2.1 – A pairs trading example

The figure shows two simulated stock prices on the left scale. In addition a dummy variable (right scale) indicates if a position is open (1) or closed (0). A position is opened if the spread exceeds a previously calculated entry–threshold value. The position is closed at the next crossing of the prices. In this specific example, a position is opened at $t = 55$, and later closed when the prices cross at $t = 65$. At $t = 120$ a position is again opened. This position stays open until $t = 330$. An intuitive way to understand the payoffs that would result from a trade is to think of the spread between the two stocks as a synthetic asset. When a position is opened the trader is effectively selling the spread short, speculating that it will decrease. When the stock prices later cross the value of the spread is zero. The trader then closes the position, and earns a profit equal to the value of the spread at the time the position was entered.

Since pairs trading is a relative value strategy, a framework for assessing the relative value development in a pair is essential. In the hypothetical example above the two stock price series both start at unity. This makes calculating the relative changes in values simple. At any given point in time the cumulative returns to the series are directly observable. Any return differences between the stocks are therefore easily calculated.

Obviously, in a real situation the stock prices will not be as well behaved as in this example, but instead start at values that vary widely. This makes the comparison more complicated. In addition, Do, Faff & Hamza (2006,4) point out that the raw price spread between two stocks is not expected to stay at a constant level, even if the stocks yield identical returns¹. This makes the raw price spread unsuitable as an indicator of when a position should be opened or closed. In order to overcome these issues we need to apply a transformation to the series. By transforming the price series we achieve price level independency and we are able to consistently assess the relative value development in the stocks.

2.1 Normalization of stock prices

In previous academic literature (Engelberg, Gao & Jagannathan, 2009; Gatev et al., 2006), a common transformation to achieve price level independency is to construct cumulative return indexes for the stocks. These indexes reflect the total return since the beginning of a period, adjusted for dividends and splits. The indexes are then rebased to some constant common for all stocks considered. In the literature this transformation is usually referred to as normalization of the stock prices.

Example – Normalized price series

As the concept of normalization is central to this thesis we will provide a concrete practical example of the procedure. In the example we will consider the intraday development of the two stocks Seadrill and Fred Olsen Energy on January 6th 2014. The figure on the next page shows the raw price series.

¹ To see this, think of the spread between two stocks A and B currently trading at 15 and 20 NOK respectively. The spread S at time 0 is equal to $S_0 = PA_0 - PB_0 = 15 - 20 = -5$. Now we assume a 100 % return in period 1. The spread then also doubles because $S_1 = PA_1 - PB_1 = 30 - 40 = -10$.

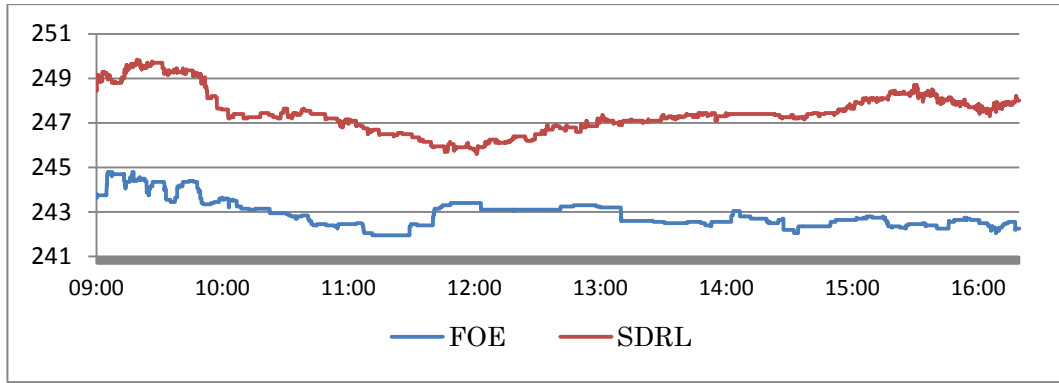


Figure 2.2 – Raw price series

We now normalize the series by rebasing both stocks to a value of one at the first observation. Mathematically this is done according to equation 2.1

$$P_{it}^N = \frac{P_{it}^{Raw}}{P_{i0}^{Raw}} \quad 2.1$$

Where P_{it}^N is the normalized price of stock i at time t and P_{it}^{Raw} the raw price of stock i at time t . The next chart shows the normalized price series.

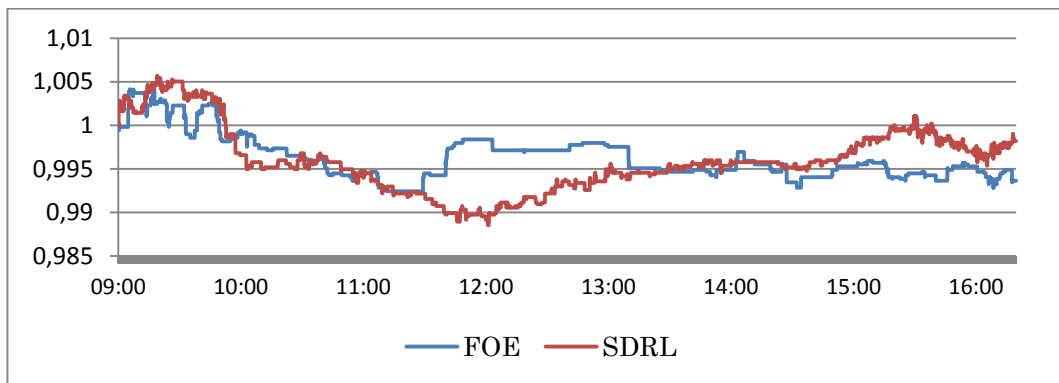


Figure 2.3 – Normalized price series

With the transformation applied the relative value development of the stocks is directly comparable. It is now possible to consistently quantify the level of divergence.

3. Pairs trading in previous literature

In this section we will review some important work previously done on the topic of pairs trading.

Gatev, Goetzman & Rouwenhorst – “Pairs Trading: Performance of a Relative Value Arbitrage Rule” (1998, 2006)

This study is one of the earliest academic investigations of pairs trading, laying the foundation for much of the subsequent research. In the paper a simple trading strategy is back-tested. Pairs are identified by finding stocks that exhibit historical co-movement in prices. Specifically, stock pairs that minimize the total distance between normalized price series are identified as potential candidates. This approach is therefore commonly referred to as the distance approach.

The dataset consists of daily closing prices from the US stock market over the period of 1962 to 2002. Significant excess returns of up to 11% annually (before costs) for self-financing pairs are reported. The authors attribute the excess returns to an unknown systematic risk factor not yet identified. They support this view by pointing out that there is a high degree of correlation between the returns to portfolios of non-overlapping pairs. The correlation is present even after accounting for common risk factors by applying an augmented² version of the Fama and French three factor model. In addition, the analysis shows that the returns to the pairs trading strategy were lower in the latter part of the sample, something the authors attribute to lower returns to the mentioned unknown risk factor. The study was first circulated as a working paper in 1998. In 2006 the sample period was extended and the paper was officially published.

²The model usually referred to as the Fama-French three-factor model was published by Eugene Fama and Kenneth French in 1992. The aim of the model is to attribute stock returns to exposure to different systematic risk factors. The three original factors included were the general market risk exposure, book-to-market ratio, and market capitalization. The augmented version used in this study includes momentum and reversal as two additional factors. For more information consult Fama & French (1992).

Vidyamurthy – “Pairs Trading, Quantitative Methods and Analysis” (2004).

In this book the author proposes the use of the cointegration framework introduced by Engle and Granger in 1987. The property of cointegration is used both to identify pairs, and to generate trade signals. Intuitively, the idea of cointegration as used in this book can be explained in the following way. Consider two series each consisting of two components: one random component and one non-random component. In addition, assume that the random component is common for both series. Then, by combining the series in a specific ratio, we obtain a new series only consisting of the non-random components. By applying the cointegration framework Vidyamurthy attempts to identify pairs of stocks where the random components cancel out. Pairs with this property would be attractive for a pairs trading strategy as their spread would be expected to stay at a constant value. The book is essentially a practitioner’s guide to pairs trading and offers no empirical results.

Elliot, Hoek & Malcom – “Pairs Trading” (2005).

In this paper the authors present an approach to pairs trading where the spread is modelled as a random variable with mean-reverting properties. Specifically, it is assumed that the spread approximately follows an Ornstein-Uhlenbeck process. The approach offers some advantages. Because the spread is modelled as a variable with certain statistical properties it is possible to forecast time to convergence and probabilities for further divergence. The paper is purely theoretical and offers no empirical analysis of the approach.

Lin, McCrae & Gulaty. –“Loss protection in pairs trading through minimum profit bounds: A cointegration approach” (2006).

The authors present a variant of the cointegration approach suggested by Vidyamurthy (2004). The cointegration coefficient (the slope in a regression between the two stocks in a pair) determines the ratio in which the stocks are bought or sold. Using this approach the necessary conditions for a trade to deliver a minimum profit is derived. The minimum profit is used to cover trading costs and profits. The empirical part of this study is limited to an analysis where one single pair of stocks is examined.

Engelberg, Gao & Jagannathan – “An Anatomy of Pairs Trading: The Role of Idiosyncratic News, Common Information and Liquidity” (2009).

Following the approach outlined by Gatev et al. (2006) the authors document significant excess returns. In addition, the paper aims to explain the factors affecting the returns. Four main findings are reported. First, the return to a trade is sensitive to the time passed between divergence and convergence. The return potential decreases exponentially with time after divergence. The authors introduce a rule where a position failing to converge in the first 10 days is closed automatically. This leads to an increase in profits from 70 bps per month to 175 bps per month. Second, it is shown that the profits to a trade are related to news affecting the companies. If the observed divergence is the result of firm specific news, the divergence is more likely to be permanent. The third observation shows that if the information shock is common to both stocks, then some of the profits to the trade can be attributed to differences in the time the market needs to adjust the prices to reflect the news. Fourth, profits are affected by owner structure and analyst coverage. If both stocks in a pair are owned by the same institutional investor the profits are reduced. Similarly, if the stocks in a pair are both covered by the same analyst, the returns are generally lower.

Do & Faff – “Does Naïve Pairs Trading Still work?” (2010)

In this study the authors attempt to replicate the results found by Gatev et al., (2006) by using the same dataset as in the original study. Their results do agree with those found in the original study with only minor discrepancies. In addition, the authors expand the data sample to include observations up to the first half of 2008. In the subsample stretching from 2003 to 2008 the excess returns have declined to a point where they are essentially zero. The authors note that there seems to be an increased risk of non-convergence in this sub-period, i.e. that the spread continues to widen after a position is opened.

Bowen, Hutchinsinon & O’Sullivan – “High Frequency Equity Pairs Trading: Transaction Costs, Speed of Execution and Patterns in Returns” (2010)

This is one of very few academic studies we have found that examines a pairs trading strategy using high frequency data. Following the approach used by

Gatev et al. (2006), the authors analyze the year of 2007 in the UK stock market. Moderate excess returns are documented. The returns are found to be highly sensitive to timing and transaction costs.

Hoel – “Statistical Arbitrage Pairs: Can Cointegration Capture Market Neutral Profits?” (2013)

Following a cointegration approach this paper back-test the performance of a pairs trading strategy over the years 2003 through 2013 in the Norwegian stock market. Hoel adopts the cointegration weighing approach proposed by Lin et al., (2006). The study shows that this implementation would have resulted in large losses, both cumulative and in most sub-periods.

George Miao – “High Frequency and Dynamic Pairs Trading Based on Statistical Arbitrage Using a Two-Stage Correlation and Cointegration Approach” (2014).

Using high frequency data from the US market Miao shows that pairs trading during 2012 and 2013 were extremely lucrative. The author reports that the strategy outperformed the S&P500 by 34 % over a 12-month trading period (before costs). The pairs formation procedure is divided in to two steps; in the first step, potential pairs are pre-selected based on their correlation coefficients. In the second step, a test for cointegration is applied to identify the best pairs. The selected pairs are then subsequently traded when deviations from the estimated relationship arise.

4. Different approaches to pairs trading

In this section we present the details of the most common pairs trading approaches.

4.1 The Cointegration approach

This approach relies on the statistical concept of stationary processes. Harris & Sollis (2003, 27) defines a time series y as stationary if the following three conditions are satisfied:

1. $E[y_t]$ is constant for all t
2. $Var[y_t]$ is constant for all t
3. $Covar[y_t, y_{t+n}]$ is constant for all t

If the spread between two assets can be confirmed to follow a stationary process it is possible that the pair can be used successfully in a pairs trading strategy. By satisfying condition 1 the expected value of the spread is constant at all times. This implies that the value of the spread is expected to revert to the mean should a deviation occur.

Next, consider two series that are both integrated of first order ($I(1)$) and therefore non-stationary³. Generally any linear combination of two such series would also be $I(1)$ and non-stationary. However, if the series share common stochastic trends it might be possible that some linear combination of the series could result in a stationary $I(0)$ series. In that case the stochastic terms cancel out and we are left with the stationary part. This concept is referred to as cointegration. A more formal definition of cointegration found in Lin et al. (2006) is quoted below.

Let $X_{1t}, X_{2t}, \dots, X_{nt}$ be a sequence of $I(1)$ time series. If there are nonzero real numbers $\beta_1, \beta_2, \dots, \beta_n$ such that $\beta_1 X_{1t} + \beta_2 X_{2t} + \dots + \beta_n X_{nt}$ becomes an $I(0)$ series, then $X_{1t}, X_{2t}, \dots, X_{nt}$ are said to be cointegrated.

This concept is essentially what we want to exploit in pairs trading. We want to identify stocks that are exposed to some set of common factors so that their relative

³ A non-stationary series is a series not meeting the conditions for a stationary process. If a non-stationary series is integrated of order 1 the series must be first differenced once in order to become a stationary $I(0)$ series. When first differencing a series the previous observation is subtracted from the current one. This yields a new series consisting of the period-to-period changes.

valuations can be reasonably well described as a fixed relationship. The prices of the stocks are then expected to follow similar paths and thus yield a stationary, and therefore mean-reverting, spread.

The standard framework for evaluating cointegration and estimating the linear relationship between stocks is based on regression analysis. The regression takes the form of equation 4.1. As discussed in section 2.1 the raw spread between two stocks is unsuitable as an indicator of the relative value development. Recall that the spread between two stocks is not expected to stay at a constant level even if the stock returns are identical. In the cointegration framework, this problem is addressed by using the natural logarithm of the prices instead of raw prices. This transformation ensures price level independency. See appendix A for further details on the transformation.

$$\ln(P_{At}) = \mu + \gamma * \ln(P_{Bt}) + \varepsilon_t \quad 4.1$$

The slope coefficient γ is referred to as the *cointegration coefficient* between the two securities. In economic terms, γ is the expected percentage increase in the price of stock A when the price of stock B increases with one percent. This translates to the expected return in stock A over some period, given the return in stock B over the same period. Vidyamurthy (2004, 106) argues that μ should be interpreted as a premium that the investor receives for holding one unit of stock A instead of γ units of stock B. It would also be possible to interpret μ in a purely technical sense without any economical meaning. A third option is to run a regression with no intercept.

After obtaining the coefficient estimate the spread between the two securities is defined as

$$S_t = \mu + \hat{\gamma} * \ln(P_{Bt}) - \ln(P_{At}) \quad 4.2$$

Given that the estimated relationship in 4.1 is valid, the variable S_t will be a stationary zero mean random variable. Notice that S_t in equation 4.2 is equal to ε_t in equation 4.1.

Practically, we can test for cointegration by analyzing the residuals resulting from the regression described in equation 4.1. The residuals are tested for stationarity by using an appropriate test such as the Dickey–Fuller test. In the Dickey–Fuller test we attempt to describe the time series as an $AR(1)$ process⁴ and then test for a unit root.

$$\varepsilon_t = \rho\varepsilon_{t-1} + v_t \quad 4.3$$

If $|\rho| < 1$ in equation 4.3 and the results are significant⁵, we conclude that the series is stationary.

The specific test of cointegration outlined here is usually referred to as the Engle–Granger two–step approach, and is widely used to determine cointegration between two variables. While the Engle–Granger procedure is simple and seems to be preferred in previous literature, other tests for stationarity are possible. Vidyamurthy (2004) analyzes the resulting series of residuals obtained from the regression in equation 4.1. The number of times the series crosses the mean are measured. A high number of crossings are interpreted as evidence for stationarity.

If the results from the stationarity test indicate cointegration the pair is selected for trading. In this step we monitor the value of the spread S_t . Any value except zero indicates a departure from the relationship estimated in 4.1. If the deviations exceed some threshold value q a position is opened. If $S_t < -q$ then, according to our estimates, stock A is overvalued compared to stock B. The trader then opens a short position in stock A and a long position in stock B. Conversely, if $S_t > q$ then

⁴ An $AR(1)$ process is a process where the current value of the series is dependent on the previous value.

$y_t = \rho y_{t-1} + \varepsilon_t$. If $\rho = 1$ we have that $\Delta y_t = \varepsilon_t$ and the series is a random walk.

⁵ The standard critical values do not apply in the DF test. Instead one must use custom critical values valid for use with the test.

stock A is relatively undervalued compared to stock B and the opposite positions are entered. The positions are closed when S_t decreases to a value lower than some threshold value.

Vidyamurthy (2004, 75) makes an attempt to link the cointegration method to the arbitrage pricing theory (APT)⁶. It is argued that the cointegration coefficient should be interpreted as the relative risk factor exposure in the two stocks. So that one unit of stock A exposes the investor to the same amount of systematic risk as γ units of stock B. Do et al., (2006, 6) criticizes this argument by pointing out that it does not account for the risk free rate of return in a way consistent with the APT. Specifically, according to the APT an investor holding γ units of stock B will receive γ units of the risk free return in addition to any return due to systematic risk exposure. On the other hand, an investor holding one unit of stock A will receive one unit of the risk free return in addition to the return due to systematic risk exposure. Equations 4.4 and 4.5 illustrate the problem.

Assume that

$$\begin{aligned} R_{At} &= R_{ft} + \gamma(\beta_1 rF_{1t} + \beta_2 rF_{2t} + \dots \beta_n rF_{nt}) \\ R_{Bt} &= R_{ft} + (\beta_1 rF_{1t} + \beta_2 rF_{2t} + \dots \beta_n rF_{nt}) \end{aligned} \tag{4.4}$$

Where rF_{it} is the excess return to the systematic risk factors, β_i the sensitivity to those factors and R_{ft} the risk free return. We now compare a position equal to one unit of stock A to a position consisting of γ units of stock B.

$$\begin{aligned} \text{Return Pos A.} &= R_{At} = R_{ft} + \gamma(\beta_1 rF_{1t} + \beta_2 rF_{2t} + \dots \beta_n rF_{nt}) \\ \text{Return Pos B.} &= \gamma R_{Bt} = \gamma R_{ft} + \gamma(\beta_1 rF_{1t} + \beta_2 rF_{2t} + \dots \beta_n rF_{nt}) \end{aligned} \tag{4.5}$$

We can see that the return to position A will differ from the return to position B due to the difference in the risk free returns components.

⁶ Stephen Ross introduced the APT in 1976 and argues that stocks are exposed to various systematic risk factors, and that the development in these factors dictates the returns to individual stocks. It then follows that stocks with identical factor exposures should have identical returns. If this is not the case then arbitrageurs would exploit this and thus eliminate the deviation. For more information on the APT we suggest that the reader consult the original paper by Ross (1976).

In the light of the discussion above, we must point out that the cointegration approach is not the only pairs trading method with little support in current asset pricing models. Sparse support in such models is common for all existing pairs trading methods. If a pairs trading strategy is able to generate excess returns this could be an indication that the current asset pricing models fail to capture all sources of systematic risk.

4.2 The Distance approach

The distance approach is the most commonly used method in previous academic literature. Gatev et al., who introduced this method in their 1998 study, explain that it is based on conversations with traders actively applying a pairs trading strategy.

Pairs are identified by calculating the sum of squared differences between normalized stock prices over some time period. The pairs are then ranked in descending order, based on their sum of squared differences. The procedure for calculating the sum of squared differences is shown in equation 4.6.

$$D = \sum_{t=1}^T (N_t^A - N_t^B)^2 \quad 4.6$$

Where D is the cumulative sum of squared differences between the normalized prices. N^i refers to the normalized stock prices. The pairs with the lowest sums will be the pairs with the highest degree of comovement and thus be the pairs with greatest potential for use in a pairs trading strategy.

A property of this approach is the implicit assumption of return parity; I.e. this method matches stocks that yield the same return in the same period. This point is sometimes mentioned as a weakness of this method (Do et al., 2006). On the other hand the authors of the mentioned study point out that the nonparametric nature of this approach leaves less room for estimation errors than more complex methods.

It is important to note that while cointegration is not explicitly tested in equation 4.6, the distance approach also relies on the cointegration property. (Gatev et al., 2006) argues that most, if not all, high-potential pairs identified, will be pairs of

cointegrated stocks. Theoretical justification for this assertion is found by assuming a pricing framework where asset prices are driven by the development in common non-stationary factors. Bossaerts & Green (1987) and Jagannathan & Viswanathan (1988) are cited as examples of such pricing frameworks. The pairs with the lowest sums of squared differences are expected to be pairs with near-equal exposure to the same systematic factors. The pairs will therefore move together in a fashion that leads to cointegration.

Assuming that an attractive pair is identified, we shift to the trading period. In this step the spread between the securities is calculated and monitored continuously. If the spread S exceeds some predefined value q a position is opened.

$$\begin{aligned} S_t &= S_{At} - S_{Bt} \\ \text{If } S_t > q &\rightarrow \text{Short position in A, long position in B} \\ \text{If } S_t < -q &\rightarrow \text{Short position in B, long position in A} \end{aligned} \tag{4.7}$$

The position is closed when the spread converges to a value equal to a predetermined closing condition. In previous literature a position is often opened when the spread deviate by more than two standard deviations as measured over the formation period. (Gatev et al., 2006; Do & Faff, 2010). In the mentioned studies the position is closed at the next crossing of the normalized price series. Naturally, a higher threshold for entering a position would yield a higher profit per trade than a lower value. On the other hand, a lower threshold-value will lead to more trades, potentially increasing the total profits. It is therefore difficult to determine whether total profits increase or decrease with higher threshold values.

4.3 The Stochastic approach

In this framework the spread between two stocks is modelled as a stochastic variable with mean reverting properties.

Previous literature does not offer any guidelines describing how to identify potential pairs. Instead it is assumed that such a pair is already identified. The

pairs identification could be done qualitatively, by choosing stocks with similar fundamental characteristics. Alternatively, one of the two previously discussed formation methods (the distance approach and the cointegration method) could be used.

Assuming that a pair of appropriate stocks is identified we provide an outline of the method. The approach assumes a continuous time framework where the spread is modelled as an Ornstein–Uhlenbeck process. The Ornstein–Uhlenbeck process is often described as the continuous time counterpart to the discrete $AR(1)$ process⁷ (Neumaier & Schneider, 1998). The spread is defined as the difference in the prices⁸ between stock A and stock B. This difference is assumed to be driven by a state variable x that follow a state process described in equation 4.8.

$$x_{k+1} - x_k = (a - bx_k)\tau + \sigma\sqrt{\tau}\epsilon_{k+1} \quad 4.8$$

Where a, b and σ are constants, and ϵ_k is a Gaussian noise term with zero mean and a standard deviation of one.

x_k will then be a normally distributed variable with the following properties

$$x_k \sim N(\mu_k, \sigma_k^2) \quad 4.10$$

With

$$\mu_k \rightarrow \frac{a}{b} \quad \text{when } k \rightarrow \infty \quad 4.11$$

And

$$\sigma_k^2 \rightarrow \frac{\sigma^2\tau}{1-(1-b\tau)^2} \quad \text{when } k \rightarrow \infty \quad 4.12$$

If the value of $\tau > 0$ is such that $|1 - b\tau| < 1$. Equation 4.8 can also be represented as

⁷ For a description of the $AR(1)$ process, see footnote 4.

⁸ In a review of the method Do et al., (2006) stresses the importance of using log-transformed prices in order to achieve a price-level independent spread.

$$x_{k+1} = A + Bx_k + C\varepsilon_{k+1} \quad 4.13$$

Where $A = a\tau \geq 0$, $B = 1 - b\tau < 1$ and $C = \sigma\sqrt{\tau}$.

We then let $x_k \cong X(k\tau)$ with $\{X(t) | t \geq 0\}$ satisfying the stochastic differential equation

$$dX(t) = (a - bX(t))dt + \sigma dW(t) \quad 4.14$$

where $W(t)$ is a standard Brownian motion. The parameters of the Ornstein–Uhlenbeck process (A , B and C in equation 4.13) can now be estimated. The parameters can easily be estimated using OLS (an example is provided in appendix B). However, the previous literature also employs more sophisticated iterative algorithms such as the Kalman Filter⁹. The variable x_{t+1} is then normally distributed with an expectation conditional on the previous observation. Appendix C shows the calculation of the conditional expectation and the conditional probability distribution.

The actual and observable spread y is then assumed to be equal to the state variable plus some noise.

$$y_k = x_k + D\omega_k \quad 4.15$$

Where $D > 0$ and ω_k is $N\sim(0,1)$. The spread is therefore also normally distributed with the same expectation as x_k . The trader can now compute the conditional expectation of y_k given y_{k-1} . A position is opened if the current spread deviates significantly from the estimated value.

Do et al. (2006) point out that this method offer several advantages. Because the spread variable is assumed to be normally distributed the variable captures the property of mean reversion explicitly. Furthermore, it facilitates forecasting, allowing the trader to calculate, amongst other things, the estimated time until

⁹ An iterative algorithm taking a series of noisy measurements observed over time as input and returns an estimate of the true value. The estimate at time t is a weighted average of the prediction given the observation at $t-1$ and the actual observation at time t . The method is named after Rudolf Kalman who developed the procedure.

convergence. On the other hand, the mentioned authors also point out that this method, like the distance approach, assumes return parity. A perhaps more fundamental involves the implicit assumption of normality in the spread between two equities. This assumption does not hold empirically (Steele, 2014).

Despite the mentioned advantages, the stochastic approach has not been tested much empirically. We have not been able to find papers empirically testing the stochastic approach.

5. Simulation testing and choice of approach

In this thesis we will focus on the distance and cointegration approaches. The methods both offer a complete strategy for pairs trading, including an algorithm for pairs selection. Furthermore, both approaches have been tested empirically and have been found capable of delivering excess returns (Gatev et al. 2006; Miao, 2014; Bowen et al., 2010).

In this part of the thesis we generate simulated data. We use this data to test the performance of the distance method versus the cointegration approach. Based on the results from the testing we select the approach to be used in the empirical part of this paper. The tests we conduct will focus only on the procedures for pairs selection. The trading step of a pairs trading strategy requires us to specify several different parameters¹⁰. Determining the superior approach is then difficult as each parameter configuration will yield different, and perhaps contradictory, results. On the other hand, the pairs selection algorithms proposed in the previous literature yield unambiguous results without requiring us to specify any parameter values. We therefore base our choice of method by examining the performance of the two selection procedures.

¹⁰ Parameters are amongst others: entry and exit thresholds, relative weight of capital placed in the long and short leg of a portfolio, number of pairs traded etc.

We generate simulated data using two different models. Common for both models is that the generated simulated stock pairs are moving according to a time-invariant relationship. The stock pairs generated will therefore mimic real world stock pairs that would be suitable for pairs trading.

In the first model we produce pairs of stocks where the cumulative returns are in parity. (I.e. if stock A increases by 1% over some time period this is also expected to be the increase in stock B). In the second model we allow for some pairs to have returns that are not in parity.

Both methods perform satisfactory under the first setup with the distance approach slightly outperforming the cointegration method. Using data generated with the second model we find that the distance approach still performs well but slightly less so than the cointegration approach. However, the most important insight resulting from the simulation testing is not directly related to the pairs ranking. The results show that the cointegration coefficient estimates appear to be very sensitive to noisy data. The estimates quickly deteriorate as the level of noise increases. Based on the results found in this part we decide to use the distance approach in the empirical part of this thesis. In the following section we present the procedure for the simulation testing and discuss the results in detail. The python code used to implement the pairs identification procedures is found in appendix D.

5.1 Model one – The granger representation theorem

The data used in the first part of the simulation testing is generated using a set of equations commonly referred to as the Granger representation theorem.

We have

$$\begin{aligned}x_t &= x_{t-1} + a_x(y_{t-1} - \gamma x_{t-1}) + \varepsilon_{xt} \\y_t &= y_{t-1} + a_y(y_{t-1} - \gamma x_{t-1}) + \varepsilon_{yt}\end{aligned}\tag{5.1}$$

Where a_y and a_x are the rates of error correction, specifying at which rate the series return to the equilibrium after deviations occur. γ specifies the equilibrium

distance between the two series. For simplicity we set $\gamma = 1$ and $y_0 = x_0 = 100$ meaning that the equilibrium distance between the series is zero. Furthermore, a_y is always equal to $-a_x$. The error terms are two normally distributed random variables with zero mean and a standard deviation of one. We examine different specifications for the error terms. In the base case the terms are uncorrelated. In the alternative cases we specify various levels of correlation between the terms. All series generated will consist of an arbitrarily selected number of 5 000 observations.

We define ten groups each consisting of ten simulated pairs. The groups are distinguished by only containing pairs with a specific value of a_x . The sets of error terms are common for all groups. This means that the first pair in the first group will use the same set of error terms as the first pair in the second group and so on. Therefore, the first pair in any group will be identical to the first pair in any other group except for the different values of a_x . The same is true for the second, third etc. pairs in each group. This is important as it ensures a consistent ranking by isolating the impact of changing the error correction rates.

Table 5.1 – Simulation parameter values

Group	a_x	No. of pairs in group
1	0.001	10
2	0.005	10
3	0.01	10
4	0.02	10
5	0.05	10
6	0.1	10
7	0.2	10
8	0.3	10
9	0.4	10
10	0.5	10

Notes: This table presents the parameter configuration for the tests on data generated by the modified granger representation model. We test the methods for pairs formation by their ability to assign the pairs to their respective groups.

a_x Error correction rate. Determines the magnitude of the error correction that follows deviations from the equilibrium between the series. Note that $a_y = -a_x$.

We will vary the error correction rate a_x according to the table above. Recall that $a_y = -a_x$ so we are effectively varying both a_x and a_y . The maximum value we

specify for the error correction rates 0.5. This value results in the closest possible relationship¹¹ between the stocks in a pair. Values lower or higher than 0.5 leads to under- or over-corrections compared to the case where $a_x = 0.5$.

The outlined setup results in a total of 100 pairs. We apply the pairs formation algorithms to the simulated pairs and observe which pairs that are identified as having the highest potential for pairs trading. Pairs with low values for a_x is expected to be placed at the bottom of the ranking. The reason for this is that the series in a pair will only loosely follow each other when the error correction rates are low. In contrast, pairs with higher values for the error correction rates tend to follow each other closely and should be ranked higher. Therefore, applying the pairs ranking algorithms to the 100 simulated pairs should sort all pairs in descending order based on their error correction rate values. As we test ten different error correction rate values, this translates to assigning the pairs to ten different groups

5.1.1 Results

The complete ranking of the pairs is found in Appendix E. The distance approach places all 100 pairs in the expected groups with no exceptions. The cointegration approach also assign all pairs to their respective groups when regressing series y on series x . However, reversing the variable ordering and regressing x on y turns out to misplace two pairs. This observation exposes an undesirable property of the cointegration approach; the ranking is sensitive to the ordering of the variables in the regression setup. In other words, regressing series y on series x gives a different test-statistic and cointegration coefficient than regressing x on y .

The above discussed problem of order sensitivity is previously discussed in Hoel (2013) and in Gregory (2011). We note that in the first paper the cointegration coefficient is used as the hedging factor. i.e. how much to go long/short in each stock in a pair. Hoel points out that if the OLS regression is used to determine the

¹¹ This argument is based on simulation results. We run 1 000 000 trials changing the error correction rate randomly between 0 and 1 and found 0.5 to be the optimal value. Code and results are available on request.

coefficients, the resulting hedge factors¹² would be inconsistent. The solution suggested in the two mentioned papers is to replace the OLS regression procedure with a different procedure called orthogonal regression. The OLS yields an expression that minimizes the vertical distance to the fitted line. Using the orthogonal regression the *perpendicular* distance from the data–points to the fitted line is minimized. This results in invertible coefficients, and yield test statistics that are insensitive to variable ordering. A more elaborate description of the orthogonal regression procedure is found in Appendix F.

When using the orthogonal regression 98 out of 100 pairs are assigned to the expected groups. The same pairs that were “misplaced” when using OLS is misplaced also when applying the orthogonal regression.

In addition to the test setup described above, we explored several cases where the error terms in equation 5.1 were correlated. These cases are important to investigate as pairs formed from related stocks are likely to experience correlated idiosyncratic shocks. Consider the two soft drink producers Pepsi and Coca Cola. It is plausible that a negative event affecting Coca Cola will result in higher sales for Pepsi. An example of such a scenario could be a case where Coca Cola experiences sudden delivery problems. Obviously this would negatively impact the revenues of Coca Cola. At the same time it is possible that some consumers originally planning to buy Coca Cola instead buys Pepsi and therefore contribute positively to the revenues of Pepsi. We model such a relationship by allowing for correlated error terms. We found no significant changes when such correlations (both positive and negative) were specified. Both methods produced rankings identical to the case with uncorrelated error terms.

5.2 Model two – The Stock & Watson Common trends model

Looking at the results from the previous analysis it becomes apparent that the Granger representation theorem produces pairs of simulated stocks with parity in the returns. This conclusion is motivated by the obtained cointegration coefficient

¹² In this setting this means that the positions taken in each stock a pair would be different in the case where the trader regress stock x on stock y compared to the case where stock y is regressed on stock x .

estimates (appendix E). The cointegration coefficient values are essentially one for the majority of the pairs. The interpretation of the slope variable in a log–log regression is the change expected in y conditional on the change in x . A cointegration coefficient equal to one therefore implies that the cumulative returns of the simulated stocks are in parity. This could be a problem for the validity of our results. Do et al. (2006, 4) criticizes the distance approach for implicitly assuming return parity between the stocks in a pair. It is claimed that this assumption is a serious limitation of the method and that the distance approach only is able to identify pairs with parity in the returns. Our setup, producing only pairs with parity in returns, might therefore unintentionally be biased in favor of the distance approach. For that reason it is necessary to explore how the methods fare when we allow for pairs to have returns related in other ways than in a one to one ratio. For this purpose we will use a modified version of a model commonly referred to as the Stock & Watson common trends model (Vidyamurthy 2004).

The original Stock & Watson model is shown in equation 5.2. The series both consist of a common nonstationary component n and an individual stationary component θ_x and θ_y .

$$\begin{aligned}x_t &= \delta_x n_t + \theta_{xt} \\y_t &= \delta_y n_t + \theta_{yt}\end{aligned}\tag{5.2}$$

The series share the common trend n_t but their exposure to the trend vary depending on the values of δ_x and δ_y . This implies that it is possible to construct a new series S by forming a linear combination of x and y such that the common non–stationary component cancels out. This would leave us with a series consisting only of the stationary components.

$$\begin{aligned}S_t = x_t - \gamma y_t &= (\delta_x n_t + \theta_{xt}) - \gamma(\delta_y n_t + \theta_{yt}) \\&= \delta_x n_t - \gamma \delta_y n_t + \theta_{xt} - \gamma \theta_{yt}\end{aligned}\tag{5.3}$$

In order for the common trend to cancel out we must have that

$$\begin{aligned} \delta_x n_t - \gamma \delta_y n_t &= 0 \\ \Rightarrow \gamma &= \frac{\delta_x n_t}{\delta_y n_t} = \frac{\delta_x}{\delta_y} \end{aligned} \tag{5.4}$$

By setting γ equal to the value found in 5.4, we are left with only the components.

$$S_t = \theta_{xt} - \theta_{yt} \tag{5.5}$$

Recall that a stationary series has the property of mean reversion. Two stocks that can be combined to produce a stationary spread therefore have the potential to be used for pairs trading.

As previously mentioned we will use a slightly modified version of the Stock & Watson model. The modifications enable us to control how closely the stock returns are related. The model we use is specified in equation 5.6.

$$\begin{aligned} x_t &= \delta_x n_t + \theta_{xt} + D_x N_{xt} \\ y_t &= \delta_y n_t + \theta_{yt} + D_y N_{yt} \end{aligned} \tag{5.6}$$

Where N_i is a non-stationary trend that is specific for each series. D_i is a constant. As in the original model n_t is the common non-stationary trend. The stationary component of the series is equal to a constant plus some Gaussian noise: $\theta_{xi} = C_i + \varepsilon_{it}$. This model enables us to control the return relationship between the simulated stocks by adjusting δ_x and δ_y . In addition we can control the strength of the relationship between the stocks by increasing or decreasing the sensitivity to the non-stationary factors. This is done by specifying different values for D_x and D_y . As in the previous test we will create a total of 100 pairs with 5 000 observations in each series. The table below shows the parameter configuration setup used to generate the data. The parameters not specified in the table are discussed separately.

Table 5.2 – Simulation parameter values

δ_x	δ_y	$D_x = D_y$	Group	Number of pairs
0.1	1	0.1	1	5
		0.5	2	5
		1	3	5
		2	4	5
		3	5	5
0.25	1	0.1	6	5
		0.5	7	5
		1	8	5
		2	9	5
0.5	1	3	10	5
		0.1	11	5
		0.5	12	5
		1	13	5
1	1	2	14	5
		3	15	5
		0.1	16	5
		0.5	17	5
1	1	1	18	5
		2	19	5
		3	20	5

Notes: This table presents the parameter configuration for the tests on data generated by the modified Stock & Watson common trends model. We test the methods for pairs formation by their ability to rank the pairs according to the setup in this table.

δ_i Sensitivity to the non-stationary factor common for both simulated stocks
 D_i Sensitivity to the non-stationary factor specific for the simulated stock.

As seen in the table above, the sensitivity to the common factor in series y is always set to one ($\delta_y = 1$). Therefore, as an example, if we set $\delta_x = 0.5$ and run a regression on the form of equation 4.1¹³, we expect the estimated value of γ to be close to 0.5. Likewise, we expect $\gamma \approx 2$ if we reverse the order of the regression.

The non-stationary terms n_t and N_t are the cumulative sums of a series of random variables that are $N \sim (0,1)$. Finally we set $C_x = C_y = 1\,000$ and $\varepsilon_x = \varepsilon_y = N \sim (0,10)$. This means that the stationary component for both series oscillates around a value of 1 000¹⁴. The second and third components of the series are the random walk elements introduced by the two non-stationary terms. Five sets of noise are generated and used in all groups. As discussed earlier this is helpful when trying to isolate the impact resulting from changes in the parameters. This setup will generate pairs with close relationships when the exposure to the individual non-stationary factors D_x and D_y are low.

¹³ That is $\ln(y_t) = \mu + \gamma \ln(x_t) + v_t$

¹⁴ The value 1 000 is selected in order to avoid negative observations.

5.2.1 Results

The cointegration approach correctly identifies all the pairs least sensitive to the specific non-stationary factors i.e. the most closely related series. However, as the amount of noise added to the series increases we observe some deviations. Generally, the approach seems capable of identifying pairs, even when the returns are not in parity. Observing the results from the distance approach we find that pairs with $\delta_x = \delta_y$ and low sensitivity to the non-stationary terms are ranked as the most promising candidates for trading. This is not surprising; recall the criticism put forward by Do et al. (2006). The way the ranking algorithm in the distance approach is constructed leads it to rank pairs with parity in the returns higher than pairs with no such parity. However, in spite of the parity assumption the distance measure also returns a decent ranking of the pairs not in parity. As in the previous case, specifying correlations between the error terms does not change the results. The complete ranking is found in appendix G.

5.3 Consequences of inaccurate cointegration coefficient estimates

Analyzing the results from the tests we notice an unexpected property of the cointegration coefficient estimates. The cointegration approach produces good estimates of the coefficient when the level of noise added to the series is low. However, as the noise level increases the estimates quickly deteriorate. When we examine the bottom half of the ranking tables, (appendix E and G) we see that several of the coefficient estimates are very poor. The most extreme cases are observed when applying the orthogonal regression; many of the estimates have the wrong sign. In addition, some of the estimated coefficients have two-digit values and are clearly absurd. The results using OLS is somewhat better but a significant fraction of the estimates still have the wrong sign and are generally inaccurate¹⁵.

When practically implementing a pairs trading strategy the trader would use financial time series to determine the relationship between the stocks in a pair.

¹⁵ It is important to notice that the inaccuracies are observed even if the Dickey Fuller tests show that the series are highly cointegrated with test statistic values below -20.

Even in the case where two stocks are identically exposed to the same risk–factors it is reasonable to assume that the observed relationship between the stocks would be a noisy realization of the underlying true relationship. This is due to random idiosyncratic shocks¹⁶ affecting stocks. As our results show, noisy data might result in inaccurate estimates of the cointegration coefficient.

When basing a pairs trading strategy on the cointegration approach the cointegration coefficient estimate is crucial. The estimate is used in all steps of the process. A signal instructing the trader to open a position is generated when the value of the spread exceeds the threshold value. The amount of capital placed in stock A relative to stock B will be dictated by the estimated coefficient value. Finally, the position is closed when the spread decreases to a level below the exit threshold. Imagine what would happen should the estimated coefficient be incorrect: Entering and exiting positions would be dictated by a false or inaccurate relationship. In addition the ratio of which the stocks are bought and sold would also be determined by an invalid relationship. It is easy to imagine that this would lead to unexpected results, and perhaps significant losses.

5.4 Choice of method

Given the ranking results produced by the two methods and the observed inaccuracies of the coefficient estimates, we decide to use the distance approach for the empirical analysis. This has two potential consequences; at one hand we might exclude possible profitable trading opportunities because we risk overlooking pairs without parity in the returns. On the other hand, given that the distance approach requires less parameter estimates, we reduce the risk of losses resulting from inaccurate estimates.

¹⁶ This could be market shocks such as liquidity shocks or business related shocks such as fires, equipment malfunction etc.

6 Testing a high frequency pairs trading strategy

In the last part of this thesis we back-test a pairs trading by strategy applying the distance approach described in the theoretical part. We first discuss the setup of the test before presenting the results.

6.1 Data

Our dataset consists of three months of high frequency intraday data. The sample period starts at January 2nd 2014 and ends on March 31st. During this period there were a total of 63 business days. The universe of stocks considered are the stocks listed on the OBX¹⁷ index. The OBX index consists of the largest 25 stocks listed on the Oslo Stock Exchange. This results in a total of 276 possible pairs. Considering only the largest stocks ensures that the stocks we use will have an adequate level of liquidity. The liquidity of the stocks selected for trading is a critical factor when testing any strategy that involves selling shares short. In low liquidity stocks there are often a low supply of shares made available for shorting. This could make a seemingly profitable opportunity impossible to exploit.

All data were downloaded daily from the Norwegian web broker Netfonds through an automated process. The data were manually adjusted for dividends and corporate actions (see appendix H for a complete list of the adjustments made). The raw data is listed in a chronological Tick-by-Tick format. The nature of the lists is such that whenever a change in the price quotes occurs, a new entry is appended. Each entry consists of a timestamp, the bid price and the ask price. The uneven update frequency of the lists leads to two practical problems. First, the

¹⁷. One stock – pharmaceutical company Algeta – is excluded from our universe. The background for this exclusion is an offer to buy all shares in Algeta for 362 NOK per share put forward the 19th of December 2013. This offer effectively pegged the price of Algeta stock to a small range just below the bid price. The trade volume also dropped significantly after this bid. This makes the stock unsuitable for pairs trading as the stock no longer is affected by factors common with other stocks but only reacts to news regarding the transaction. Due to the low trade volume in it is also unclear whether it would have been possible to short Algeta during this period. We therefore argue that a trader following a Pairs Trading strategy would have excluded Algeta from the possible pairs based on information available at the time. The bid was accepted in February 2014 and Algeta was eventually delisted in the beginning of March 2014.

price observations of the various stocks do not coincide in time; there can easily be a change in one stock at one time without a corresponding change in any other stock. Second, the number of observations will vary from stock to stock as there is a different number of price changes occurring in different stocks. Both these problems are however easy to overcome. The lists contain all changes happening over the trading day. This implies that if no change is recorded at a given point in time, the bid and ask prices offered must still equal to the prices in the previous entry. It is therefore possible to compare the observations and “stretch” out the lists by copying the last observation. This ensures that all stocks will have the same number of observations, and that the observations coincide in time. We illustrate the expansion procedure with a concrete example.

Example – Expanding lists of observations

Stock A			Stock B		
Time (HHMMSS)	Bid	Ask	Time (HHMMSS)	Bid	Ask
090000	100	100.1	090000	51	51.5
090002	100	100.2	090001	51.1	51.4
090003	99.9	100.1	090004	51.1	51.2
...

Lists with raw observations

In this example we have lists of observations of two fictional stocks. As we can see, the timing of the observations is not equal. While both stocks have an observation at time 09:00:00, only stock B has an observation one second later. Due to the fact that the lists are updated only when changes in quotes occur we can easily solve this problem. The information at time 09:00:01 in stock A must be equal to the quote at 09:00:00. We therefore create a new observation at 09:00:01 with identical values as the 09:00:00 observation. This process is repeated for both stocks and for all timestamps. The tables below show the lists of observations that would result when applying the process to the lists in this example.

Stock A			Stock B		
Time (HHMMSS)	Bid	Ask	Time (HHMMSS)	Bid	Ask
090000	100	100.1	090000	51	51.5
090001	100	100.1	090001	51.1	51.4
090002	100	100.2	090002	51.1	51.4
090003	99.9	100.1	090003	51.1	51.4
090004	99.9	100.1	090004	51.1	51.2
...

Lists with observations after completion of the expansion procedure.

After the expansion procedure is completed for all stocks we trim the lists by keeping only the last observation in a given second. This gives us 26 400 observations for each stock each day, equally spaced at one second intervals. Trimming the lists has two benefits: First, it significantly reduces the time needed for computation. Second, it enables us to analyze the impact on the returns if we enforce a delay between trade signal and order execution.

6.2 Formation and trading periods

We specify a length of the formation period that is twice the length of the trading period, as this seems to be the standard in the academic literature (Gatev et al., 2006; Do & Faff, 2010; Miao 2014). We further follow the mentioned studies and construct a series of overlapping sample periods. This is important to maintain a separation between the in and out of sample periods. The figure below shows a graphical representation of the formation and trading periods.

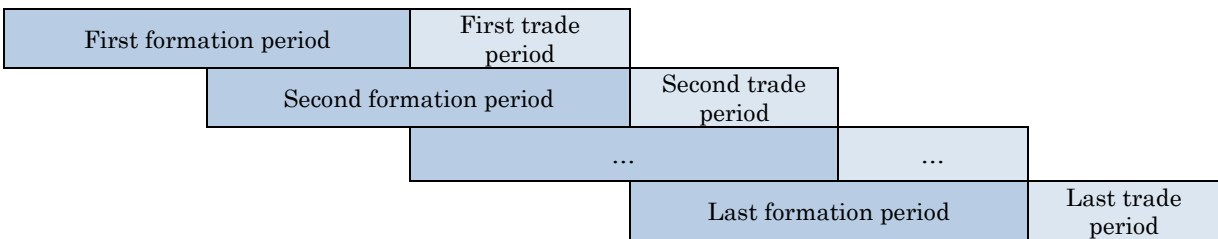


Figure 6.1 – Graphical representation of trading and formation periods.

In this thesis we test the strategy using two different lengths for the periods. In the first setup we let the formation period be two days followed by a one day trading period. In the second setup we double the timeframe and specify formation and

trading periods of four and two days respectively. Engelberg et al., (2010) show that the profits to a pairs trading strategy are related to news events. Specifically, it is shown that there is a delay between the time a news event is published and the time the stock prices fully reflect the news. The speed in which the market participants react to such news sometimes differs slightly from stock to stock. These differences cause a lead–lag relationship in the prices that could explain some of the profits to a pairs trading strategy. A two day trading period therefore allows us to capture profits arising from situations where prices diverge on one day and return to equilibrium the subsequent day.

6.3 Practical Implementation of the distance approach

6.3.1 Pairs formation

The pairs are formed by applying the distance approach. Pairs are ranked based on their total sum of squared deviations when comparing the normalized stock prices. The top n pairs are then selected for trading. A more detailed description of the procedure is available in section 4.2.

Practically we must adjust for missing quotes in the opening seconds. This is necessary because there are small inequalities as to when trading in the different stocks start. Normally quotes are available for all stocks within the first 20 seconds of the trading day. By staggering the start of the calculation until all stocks have quotes we ensure an equal number of observations for all pairs. We use the midpoint of the bid and ask prices as an approximation for the true price of the asset. The use of mid–prices greatly simplifies the practical aspect of implementing the formation algorithm.

6.3.2 Pairs Trading

With the candidate pairs identified we continue to the trading period. As in the formation period we use the midpoint between the bid and ask quotes as an estimate of the true price. We then use these prices to calculate and monitor the level of the spread. When the spread signals a trade the transaction is carried out

using the actual bid and ask prices. This is crucial as the indirect transaction costs resulting from the bid/ask spread could erode any profitable patterns found using mid-prices.

We enter a position when the spread between the two stocks exceeds the specified entry-threshold (q in equation 4.6). The position is closed at the next crossing of the normalized price series. After a crossing the pair is immediately available for trading should the prices diverge again. If a position is open at the end of the trading period the position is automatically closed. We follow Gatev et al. (2006) and let the proportions of the long and short positions be equal. That is; the market value of the long position is equal to the market value of the short position.

When a specific pair is evaluated we wait until both stocks in the pair have quotes before we start the trading. As previously mentioned all stocks are usually trading normally within 20 seconds after the exchange opens. We allow for trading on quotes posted between 09:00 and 16:19:59¹⁸. As suggested by Nath (2003) and Caldeira & Moura (2012) we implement an option to specify a stop-loss limit. This function will automatically unwind a position when a loss of a certain amount occurs. If a position is closed because the stop-loss limit was exceeded the pair cannot be reopened for the rest of the trading period.

We make an attempt to protect the strategy from losses resulting from entering irrational positions. By an irrational position we mean a position that would immediately result in a loss due to excessive costs. Specifically, we will not enter a position if the estimated loss resulting from the bid/ask spreads together with the approximate commission costs (L), are larger than the potential profit (π) that would result from convergence.

The estimated transaction costs are calculated according to equation 6.1.

$$L_t = \frac{(P_{At}^{Ask} - P_{At}^{Bid})}{2} + \frac{(P_{Bt}^{Ask} - P_{Bt}^{Bid})}{2} + 4C \quad 6.1$$

¹⁸ The opening hours for Oslo Stock Exchange are from 09:00 to 16:30. However, at 16:20 the continuous trading stops and a closing auction is initiated. During this auction existing orders are matched before trading is closed. The nature of the auction makes the bid/ask quotes posted unreliable as there is no guarantee that an order would be filled. Therefore trading after 16:20 is not permitted in our tests.

The approximate commission per trade C is multiplied¹⁹ by 4. Equation 6.2 shows the calculation of the potential profits using mid–prices.

$$\pi_t = P_{At}^{Midprice} - P_{Bt}^{Midprice} \quad 6.2$$

A position cannot be entered if $L_t > \pi_t$. When performing this check we implicitly assume that the spreads will be equal both when we enter and exit the positions. Obviously we can only observe the spreads when we enter a position. Therefore a trade could still result in a loss even if the above condition is fulfilled.

The python code for the trading routine is found in appendix I.

6.4 Return computation

As the strategy involves taking simultaneous short and long positions it is not immediately clear how profits should be calculated. The reason for this is due to the possibility of financing a significant part of the long leg with the proceeds from the short leg. This results in a gearing effect and it is unclear how much additional capital the investor needs to commit to the strategy. This complicates the return calculation.

The payoffs to a long/short pair are a series of positive cash–flows that occurs at different times during the trading period, whenever a trade is successfully completed. In addition, at the end of the period any open position is closed, either resulting in a profit or in a loss. It is also possible that no position was opened during the trading period. Obviously, in that case there would be no cash–flows from the pair.

We will follow Gatev et al. (2006) and compute the returns as the return to total committed capital. This is a conservative method that acknowledges the fact that there are certain margin requirements that need to be met when going selling stock short. Expression 6.3 shows how returns to committed capital are computed.

¹⁹ We multiply the commission per trade by four since a total roundtrip consists of two transactions when entering the position and two transactions when exiting. This is an approximation because the costs associated with exiting a position are dependent of the actual value of the position at the time of realization.

$$r_{pair} = \frac{\sum_{i=1}^n p_i}{L + \tau S} \quad 6.3$$

Where the numerator is the sum of the cash flows generated by the pair. L and S refer to the amount of capital placed in the long and short positions. τ refers to the capital requirement fraction as a percentage of the market value of the position being sold short. This fraction ranges from as low as 15 and up to 100 %, depending on the specific volatility, recent history etc. of the asset. We therefore follow Gatev et al. (2006) and Hoel (2013) and set $\tau = 1$. This will result in more conservative return estimates than had a lower value been specified.

Profits from a pair during a period are reinvested so that the sum of the cash flows is the compounded return to the pair over the period. When forming portfolios of several pairs we let the amount of capital committed to each pair be equal. The total portfolio returns are therefore simply the arithmetic average of the returns from the pairs in the portfolio.

Calculating returns as described here ensures that the returns we obtain will be comparable with the returns reported by Gatev et al. (2006) and Hoel (2013). The latter study is of special interest as it is the only paper we know of that examines pairs trading strategy in the Norwegian stock market.

6.5 An algorithmic representation of the test setup

We have now outlined all major steps in the trading strategy we will test. In order to sum up these steps we now present an algorithmic representation of the complete testing procedure.

1. Construct cumulative return indexes for the stocks over the formation period. The indexes include dividends and corporate actions. Normalize the indexes so that they all start with a value equal to one.
2. Calculate the sum of squared differences for all possible combinations of stocks. Rank the pairs based on their sum of squared differences. The n pairs with the lowest sums are selected for trading.

3. From this point on we consider a specific pair selected for trading. Normalize the prices over the trading period by following the same procedure as in step 1.
4. Monitor the spread S_t between the two stocks. If the absolute value of the spread exceeds the limit q a trading signal is given.
 - 4.1 If $S_t > q \rightarrow$ Signal short position in A , long position in B
 - 4.2 If $S_t < -q \rightarrow$ Signal short position in B , long position in A
 - 4.3 Control for costs using expression 6.1. If the estimated costs (L) are lower than the current spread (S_t) open a position in accordance with the trading signal.
5. Exit an open position if one of the following occurs:
 - 5.1 The stocks converge so that S_t crosses zero. The pair is still available for trading, return to step 4.
 - 5.2 The position is closed by the stop loss condition being met when the value of the position declines to a value below the maximum loss tolerated. The pair is no longer available for trading.
 - 5.3 The trading period is over.
6. Calculate and report return on pair. Return to step 3 if there are still pairs left to be considered.
7. Report total portfolio return (the average return to the top n pairs) for the trading period.
8. Report compounded returns for all trading periods.

6.6 Results

We now present the results of the empirical analysis. We will first consider an unrestricted case where pairs are allowed to be formed with stocks from different industry sectors. Next, we examine a restricted case where pairs are formed only by matching stocks that both operate within the same industry sector.

We report the results gross of trading costs before examining the impact of such costs in detail. In addition we will assess the impact when the trader faces a delay between trade signal and order execution.

6.6.1 Unrestricted pairs matching

The table below shows the results from the unrestricted case. The returns, standard deviation and Sharpe ratios are annualized figures. The trade column shows the number of roundtrips completed during the three month sample period.

Table 6.1 – Results Unrestricted Case

Portfolio Parameters				Top 5 Ranked Pairs				Top 10 Ranked Pairs				All Pairs			
No	SL	q	Time	Return	σ	SR	Trades	Return	σ	SR	Trades	Return	σ	SR	Trades
1	2 %	1	2:1	-60.81 %	5.4 %	-11.8	903	-56.33 %	4.2 %	-14.2	1 709	-42.68 %	3.1 %	-14.6	26 664
2	2 %	2	2:1	-45.14 %	5.8 %	-8.3	473	-39.29 %	4.6 %	-9.2	879	-24.17 %	2.5 %	-11.0	12 026
3	2 %	3	2:1	-36.87 %	5.4 %	-7.4	300	-29.28 %	4.1 %	-7.8	564	-14.56 %	1.7 %	-10.2	6 484
4	2 %	5	2:1	-26.46 %	4.2 %	-7.0	148	-20.26 %	3.3 %	-7.1	271	-6.58 %	1.0 %	-9.3	2 296
5	5 %	1	2:1	-60.47 %	5.2 %	-12.2	903	-54.31 %	4.2 %	-13.8	1 712	-42.71 %	3.4 %	-13.4	26 709
6	5 %	2	2:1	-44.48 %	5.6 %	-8.5	473	-37.34 %	4.8 %	-8.4	880	-23.62 %	2.7 %	-9.9	12 040
7	5 %	3	2:1	-35.89 %	5.1 %	-7.7	300	-26.92 %	4.3 %	-7.0	564	-14.34 %	2.0 %	-8.8	6 488
8	5 %	5	2:1	-24.26 %	3.7 %	-7.4	148	-16.40 %	3.2 %	-6.0	271	-6.62 %	1.5 %	-6.5	2 298
9	None	1	2:1	-60.47 %	5.2 %	-12.2	903	-54.31 %	4.2 %	-13.8	1 712	-43.06 %	3.5 %	-13.1	26 709
10	None	2	2:1	-44.48 %	5.6 %	-8.5	473	-37.34 %	4.8 %	-8.4	880	-23.81 %	2.8 %	-9.8	12 040
11	None	3	2:1	-35.89 %	5.1 %	-7.7	300	-26.92 %	4.3 %	-7.0	564	-14.44 %	2.0 %	-8.7	6 488
12	None	5	2:1	-17.23 %	3.7 %	-5.5	148	-16.40 %	3.2 %	-6.0	271	-6.18 %	1.4 %	-6.7	2 298
13	2 %	1	4:2	-41.24 %	5.5 %	-8.1	452	-38.85 %	3.9 %	-10.8	857	-24.86 %	2.9 %	-9.5	12 180
14	2 %	2	4:2	-30.98 %	5.5 %	-6.2	240	-27.07 %	4.5 %	-6.7	447	-12.62 %	2.1 %	-7.4	5 560
15	2 %	3	4:2	-20.76 %	4.8 %	-5.0	157	-20.76 %	3.9 %	-6.1	283	-6.83 %	1.6 %	-6.0	3 078
16	2 %	5	4:2	-21.25 %	3.3 %	-7.3	78	-17.91 %	2.8 %	-7.4	137	-2.17 %	1.1 %	-4.7	1 106
17	5 %	1	4:2	-37.00 %	5.6 %	-7.1	464	-38.32 %	4.6 %	-9.0	869	-23.31 %	3.5 %	-7.4	12 293
18	5 %	2	4:2	-29.17 %	6.3 %	-5.1	245	-26.73 %	5.5 %	-5.4	452	-11.14 %	2.7 %	-5.2	5 582
19	5 %	3	4:2	-22.46 %	5.6 %	-4.6	159	-19.60 %	4.7 %	-4.8	285	-6.22 %	2.2 %	-4.3	3 082
20	5 %	5	4:2	-15.56 %	4.2 %	-4.4	80	-13.88 %	3.2 %	-5.3	139	-3.10 %	1.6 %	-3.9	1 110
21	None	1	4:2	-37.00 %	5.6 %	-7.1	464	-38.32 %	4.6 %	-9.0	869	-23.27 %	3.5 %	-7.4	12 294
22	None	2	4:2	-28.66 %	6.2 %	-5.1	245	-26.47 %	5.4 %	-5.5	452	-11.00 %	2.7 %	-5.2	5 582
23	None	3	4:2	-21.90 %	5.5 %	-4.6	159	-19.31 %	4.6 %	-4.8	285	-5.63 %	2.1 %	-4.2	3 082
24	None	5	4:2	-15.56 %	4.2 %	-4.4	80	-13.88 %	3.2 %	-5.3	139	-1.96 %	1.3 %	-3.8	1 110

Notes: This table presents the annualized returns on a pairs trading strategy where the pairs are forced to be formed with stocks from the same industry sector.

No	Parameter identification number
SL	Stop-loss threshold
Time	Length of formation period : Length of trading period
σ	Annualized standard deviation on portfolio
q	Threshold for opening a position. Numbers of standard deviations.
SR	Annualized Sharpe ratio.
Trades	Number of roundtrips during the 3M testing period. One roundtrip translates to four trades. Two when opening a position and two when the position is closed.

We test several different configurations for the different trade parameters (the entry–threshold value q , the stop–loss threshold and the period lengths). In addition we vary the number of pairs traded. The table is interpreted in the following way: each line describes a different parameter configuration that is specified in the leftmost group of columns. The next three column groups show the results for the Top 5 Pairs, Top 10 and Pairs and All Pairs portfolios²⁰ given the specified parameter configuration.

The returns to the strategy are negative for all tested configurations of the parameters. We do however make some interesting observations. It is clear that some parameter configurations consistently yield better results than other.

First, we observe that the results notably improve when the length of the estimation and trade periods are doubled. We speculate that the main reason for this is due to premature exits from positions that would eventually have converged had they not been closed. When the trading period is only one day a large fraction of pairs are still open at the end of the trading period. In our setup any position still open at the end of the trading period is automatically closed, possibly resulting in a loss. Extending the trading period to two days allows for a number of these pairs to converge and yield a profit. Although not tested in this paper it is possible that further extending the periods could yield even better results.

The next observation involves the stop–loss threshold. Raising the stop–loss threshold appears to lead to improved results. Similarly to the first observation, this indicates that low stop–loss threshold values lead to premature exits from positions that otherwise would eventually converge. As an interesting side note, we observe that raising the threshold above 5 % has close to zero impact on the results. This indicates that only a marginal fraction of the positions results in losses exceeding 5 %.

The last observation concerns the entry threshold. Changing this parameter has a substantial impact on the returns. Specifically, higher values lead to considerable increases in the returns. There might be several explanations for this. It is possible

²⁰ Recall that in the formation period we rank all pairs based on their attractiveness for pairs trading. The Top n notation here refers to the portfolios formed by selecting the top n pairs for trading.

that our attempt to control for the impact of the bid/ask spread (equation 6.1) is inadequate. A high trade frequency could then lead to losses due to the transaction costs imposed by this spread. A different explanation could be related to the two previous observations: a low threshold value might open many positions that fail to converge before the trade period is over because they first continue to diverge. It is possible that the pairs would have converged given enough time.

In sum, the observations we make provide us with some insight as to what an optimal parameter configuration for a pairs trading strategy might look like.

6.6.2 Restricted case – Pairs from corresponding industries

We now examine the returns to the case where the components of the pairs are required to be stocks belonging to the same industry sector. Gatev et al. (2006) examine the returns to a restricted case where all stocks are assigned to one of four major industry groups as defined by Standard & Poor's. The four groups are Utilities, Financials, Transportation and Industrials. When enforcing the restrictions the authors report slightly lower annualized returns compared to the unrestricted case. On the other hand Do & Faff (2010) argue that forming portfolios of restricted pairs can be seen as a first step to implement fundamental factors in a pairs trading strategy. Do & Faff report significant improvements in the returns when a more refined classification scheme is implemented. In their study stocks are assigned to industry categories by applying a classification system introduced by Fama and French (1997). Following this system stocks are assigned to one of 48 categories.

Oslo Stock Exchange uses the GICS²¹ system to classify the listed stocks. When matching stocks in the restricted case, we specify that a pair can only be formed if the stocks have identical²² GICS identification codes. The introduction of this

²¹ The Global Industry Classification Standard is a classification system developed by the American company MSCI. The GICS system is made up of 10 sectors, 24 industry groups, 67 industries and 156 sub-industries.

²² It would also be possible to specify a less stringent restriction in which only some of the digits must be identical.

condition reduces the number of possible pairs²³ to 17. Not surprisingly a large proportion of these pairs consist of companies from within the oil & gas industry. The complete list of industry matched pairs is available in appendix J.

Table 6.2 – Returns Restricted Case

Portfolio Parameters				Top 5 Pairs				Top 10 Pairs				All Pairs			
No	SL	q	Time	Return	σ	SR	Trades	Return	S	SR	Trades	Return	S	SR	Trades
25	2 %	1	2:1	-46.30 %	5.5 %	-9.0	703	-40.35 %	5.3 %	-8.2	1244	-32.72 %	4.7 %	-7.6	1793
26	2 %	2	2:1	-19.28 %	4.9 %	-4.5	349	-17.29 %	4.5 %	-4.5	581	-10.83 %	3.5 %	-3.9	790
27	2 %	3	2:1	-13.07 %	4.4 %	-3.7	187	-8.30 %	3.5 %	-3.2	316	-4.70 %	2.7 %	-2.9	410
28	2 %	5	2:1	-3.57 %	2.7 %	-2.4	72	-4.28 %	2.3 %	-3.2	107	-1.15 %	1.8 %	-2.3	131
29	5 %	1	2:1	-46.00 %	5.6 %	-8.8	703	-38.67 %	5.2 %	-8.1	1250	-30.85 %	4.5 %	-7.5	1800
30	5 %	2	2:1	-18.95 %	5.0 %	-4.4	349	-14.20 %	4.2 %	-4.1	583	-8.25 %	3.3 %	-3.5	793
31	5 %	3	2:1	-13.24 %	4.4 %	-3.7	187	-5.15 %	3.2 %	-2.6	316	-1.14 %	2.2 %	-1.9	410
32	5 %	5	2:1	-3.07 %	2.7 %	-2.3	72	0.05 %	1.6 %	-1.9	107	2.55 %	1.3 %	-0.3	131
33	None	1	2:1	-46.00 %	5.6 %	-8.8	703	-38.67 %	5.2 %	-8.1	1250	-30.44 %	4.4 %	-7.6	1800
34	None	2	2:1	-18.95 %	5.0 %	-4.4	349	-14.20 %	4.2 %	-4.1	583	-7.58 %	3.2 %	-3.3	793
35	None	3	2:1	-13.24 %	4.4 %	-3.7	187	-5.15 %	3.2 %	-2.6	316	-1.14 %	2.2 %	-1.9	410
36	None	5	2:1	-3.07 %	2.7 %	-2.3	72	0.05 %	1.6 %	-1.9	107	2.55 %	1.3 %	-0.3	131
37	2 %	1	4:2	-25.67 %	6.6 %	-4.4	340	-27.46 %	5.6 %	-5.5	573	-24.59 %	5.1 %	-5.4	819
38	2 %	2	4:2	5.49 %	5.7 %	0.4	170	-2.37 %	5.1 %	-1.0	272	-0.59 %	3.7 %	-1.0	371
39	2 %	3	4:2	4.53 %	3.5 %	0.4	83	-1.67 %	3.4 %	-1.4	138	-1.00 %	2.4 %	-1.7	178
40	2 %	5	4:2	16.53 %	2.8 %	4.8	35	8.81 %	2.0 %	2.9	53	3.40 %	1.4 %	0.3	62
41	5 %	1	4:2	-13.22 %	5.2 %	-3.1	349	-17.02 %	5.0 %	-4.0	585	-17.66 %	4.8 %	-4.3	832
42	5 %	2	4:2	16.95 %	4.6 %	3.0	171	7.86 %	4.7 %	1.0	275	5.58 %	3.5 %	0.7	374
43	5 %	3	4:2	9.40 %	3.6 %	1.8	83	5.74 %	3.4 %	0.8	138	4.03 %	2.5 %	0.4	178
44	5 %	5	4:2	16.87 %	2.9 %	4.8	35	12.61 %	2.8 %	3.5	53	5.66 %	1.8 %	1.5	62
45	None	1	4:2	-13.22 %	5.2 %	-3.1	349	-15.81 %	5.0 %	-3.8	585	-17.54 %	0.5 %	-45.6	832
46	None	2	4:2	16.95 %	4.6 %	3.0	171	7.71 %	4.8 %	1.0	275	5.50 %	3.5 %	0.7	374
47	None	3	4:2	9.40 %	3.6 %	1.8	83	5.60 %	4.4 %	0.6	138	3.95 %	2.5 %	0.4	178
48	None	5	4:2	16.87 %	2.9 %	4.8	35	12.61 %	2.8 %	3.5	53	5.66 %	1.8 %	1.5	62

Notes: This table presents the annualized returns on a pairs trading strategy where the pairs are forced to be formed with stocks from the same industry sector.

No	Parameter identification number
SL	Stop-loss threshold
Time	Length of formation period : Length of trading period
σ	Annualized standard deviation on portfolio
q	Threshold for opening a position. Numbers of standard deviations.
SR	Annualized Sharpe ratio.
Trades	Number of roundtrips during the 3M testing period. One roundtrip translates to four trades. Two when opening a position and two when the position is closed.

²³ We refer to the previous discussion regarding the exclusion of Algeta. We point out that in this restricted setup Algeta would have been excluded in all cases. This because it was the only pharmaceutical company listed on the OBX index.

The above table presents the full set of results obtained under this setup. The results are very interesting. Through analyzing the results from the unrestricted case we identified a set of parameter configurations that were associated with better returns. The same parameterization is superior also in this case; higher entry–threshold values and stop–loss values increase returns. In addition, as in the unrestricted case, the returns are greatly improved when the formation and trade periods are extended to four and two days respectively.

The restricted case results support the observations we made when examining the unrestricted returns: the ideal set of parameters reduces the risk of entering positions that would later be prematurely exited. Even more interestingly, the combination of the favorable parameter configurations and the augmented pairs matching procedure yields returns as high as 16.95 % annually. This is an indication that a carefully designed pairs trading strategy could in fact generate positive excess returns.

We note that that as the threshold–value for entering a position is increased the standard deviations of the returns decrease. This property leads to high Sharpe ratios in the cases where the entry–threshold is set to high values. Furthermore, as one would expect, high entry–threshold values significantly reduces the trading frequency. This last observation is important as the transaction costs resulting from excessive trading could drive the profits from the strategy down. We will later examine the effects of transaction costs in detail.

*

In the unrestricted case the returns are strongly negative even when the most desirable parameter configurations are specified. Switching to the restricted case the same parameter configurations yield extremely lucrative returns, with Sharpe ratio values as high as 4.8. Given these differences in returns it is natural to raise the question of what causes these discrepancies. The answer is simply that different pairs are selected for trading as we move from the unrestricted to the restricted case. We investigate the pairs selected in both cases and find that the

overlap in the Top 5 portfolios formed is negligible²⁴. There could be multiple reasons for this. One possibility is that some of the estimated relationships are completely spurious. This happens if stocks move closely over the course of the formation just out of chance. This is to some extent supported by our data; in the unrestricted case there is little consistency in the pairs selected from period to period. In addition, a large fraction of high ranking pairs is made up by stocks that appear to be completely unrelated when comparing properties such as industry, geographical area of operation etc. Turning to the restricted case we find that some pairs are consistently identified as the most lucrative pairs and therefore included in the Top 5 Pair portfolio in several periods.

Our results are in line with those reported by Do & Faff (2010); there are substantial gains to be made when the distance approach is augmented by pre-screening the possible pairs based on industry.

6.7 Impact of transaction costs and timing constraints

The initial results suggest that it could be possible to generate significant excess returns by following a pairs trading strategy. In this section we will examine how the returns are affected when we deduct commissions and short fees. In addition, we will also investigate the impact on the returns when the trader faces timing constraints. Specifically, we investigate if the returns decline when there is a difference between the time a trade is signaled and the time the trade is executed. Obviously, the parameter configuration setups that resulted in losses before transaction costs will still do so after subtracting such costs. For that reason we will focus only on the parameter configurations that resulted in positive returns²⁵ during the previous tests.

²⁴ Specifically, we considered the selected pairs resulting from a formation period of four days. We compared the top 5 pairs in the restricted and unrestricted cases and found overlap in only 7 of 28 periods. In none of these seven periods were there more than one pair being selected in both cases.

²⁵ The configurations considered will be the Top5/10/All portfolios in the restricted case with parameter identification numbers 38 to 48 excluding number 41 and 45 in table 6.2.

6.7.1 Commissions and short fees

We will now analyze whether the positive results found are robust when controlling for trading costs. Do & Faff (2012) review a large body of literature concerning price anomalies in the stock market. In a large proportion of the reviewed studies positive returns disappear when transaction costs are controlled for. It will therefore be interesting to see if the return to the pairs trading strategy remains positive after such costs.

In the later years the commission on stock transactions has been declining. At the time of writing the most price-competitive Norwegian retail brokers offer commission at rates between 3 and 10 basis points per trade²⁶. It is plausible that an institutional investor would achieve even better rates. We will test two levels of commissions: one low level of 3 bps. and one high level of 10 bps.

In addition to commission an investor wanting to sell stock short would have to pay a fee to do so. This is the price the trader needs to pay in order to borrow the stock being sold short. In the retail market the annual short fee rates are in the range of 400 – 500 bps. In all our tests we specify a level of 500 bps. If a trade is opened and closed within one trading day we subtract shorting fees for one full trading day. Similarly, if the position is held overnight we will deduct fees as if the position were open over the entire two days. In practice it is seldom the case that the broker would charge the investor fees for intraday short positions. Therefore, our chosen implementation of short-fees could lead to some overestimation of the actual fees. It is our belief that it is better to report a more conservative level of returns than to report artificially high returns due to a failure to account for all costs.

²⁶ Netfonds.no charges a fee of 5 bps for standard costumers. For costumers whose trading volume exceeds 10M NOK monthly the rate is 3 bps. Nordnet.no charges fees ranging from 3.9 to 10 bps depending on trade volume.

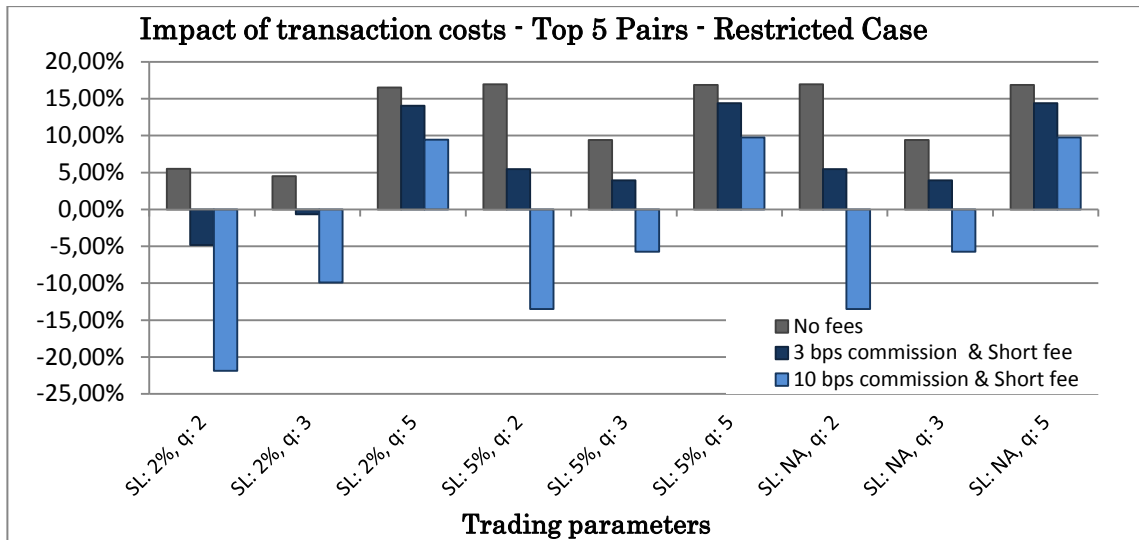


Figure 6.2 – Impact of transaction costs. SL: stop–loss threshold value, q: entry threshold.

The chart above shows the results for the top 5 pairs given the various parameter configurations. Controlling for costs significantly reduces profits irrespective of parameter values. As is to be expected the difference is most pronounced in cases with low thresholds for entering a position. The number of roundtrips made in the cases where entry threshold is lowest (two standard deviations) is around 170. In contrast, only 35 roundtrips are made in the high–threshold cases (five standard deviations). A roundtrip consists of four trades; two when a position is opened and two upon exiting. The returns in the low–threshold cases are completely eroded by the transaction costs.

Shifting to the high–threshold cases the impact of costs is also clearly visible: the profits decline by about 15 % (42 %) when comparing the annualized returns before and after 3 bps (10 bps.) are charged per transaction in addition to the short fees. Interestingly, a strategy parameterized so that it avoids excessive trading still delivers in economically significant returns after costs. For the configurations where a five standard deviation threshold is specified, the annualized returns after costs are still very lucrative at just above 14 % (3 bps. commission). The complete set of results with additional statistics is available in appendix K.

6.7.2 Speed of execution

We now turn to the issue of execution speed. So far we have assumed that a trade can be executed in the exact moment a trade is signaled. However, a trader might face constraints with respect to how fast a trade can be executed. In a study from 2010 using intraday data Bowen et al. report significant reductions in the returns when implementing a “wait-one-period” restriction²⁷. Similarly, Gatev et al. (2006) report considerably lower returns when they enforce a one day delay from the time a trade is signaled to the time of execution.

Studying the effects of execution speed is interesting for two reasons. Obviously, it is of great interest to see if the strategy tested in this paper remains profitable after implementing the restriction. However, analyzing how fast profitable opportunities disappear in a market is very interesting subject in its own right. The disappearance of such opportunities reveals something about the efficiency of the market studied. In a market with a high degree of efficiency such opportunities should be eliminated by arbitrageurs (such as traders applying a pairs trading strategy like the one we test) shortly after they come in to existence.

In this test we will investigate the impact of timing on returns by delaying the execution of a trade by 1, 10 and 60 minutes. The results for the Top 5 Pairs with different parameter configurations are presented in the figure below. Note that the results are gross of commission and other fees. The full set of results is available in appendix L.

²⁷ Their dataset consists of observations at 60 minute intervals. Their restriction therefore translates to a full 60 minute delay between trade signal and execution.

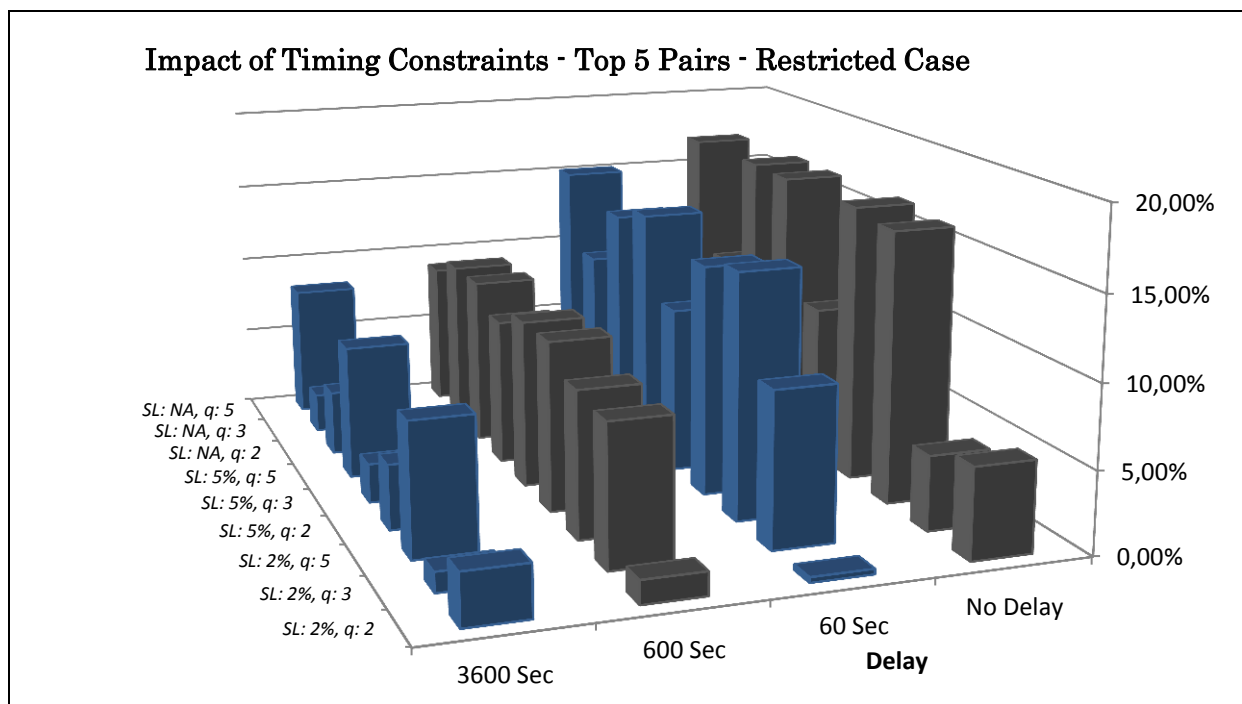


Figure 6.3 – Impact of a delay between trade signal and trade execution. SL: stop-loss threshold, q: entry threshold.

The results are in line with findings in the previous literature (Bowen et al., 2010). The returns are significantly reduced when enforcing a delay. The parameter configurations that were most profitable before any delay see the returns decline by around 50 % when a 60 minute delay is enforced. However, the strategy remains positive even when implementing a very conservative delay such as 60 minutes. Today, with the aid of computers most sophisticated traders are able to execute a trade within seconds. It is therefore the results given more moderate delays that are most interesting. The results show that strategy still remains very lucrative after specifying moderate delays of 60 or even 600 seconds.

It is interesting to see that the returns are still positive even when a delay of 60 minutes is specified. This indicates that the market participants did not fully exploit the possible profit opportunities that were present in the market during the time of our sample.

As an interesting side note, we observe that the returns to the portfolios with a low entry-threshold actually increase when specifying moderate delays. We speculate that this is because a significant share of the positions entered at such a low

threshold continues to diverge after entry. If the stocks continue their trends in the delay period the divergence would be larger when we eventually enter the position. This would lead to increased profits. In other words: a setup with a low entry threshold combined with a enforcing a delay would be similar to a setup with a higher threshold and no delay.

6.7.3 Concluding remarks regarding commissions and execution speed

We have shown that pairs trading remains positive even after controlling for transaction costs and enforcing restrictions to the speed of order execution. However, we have examined at the impact of costs and timing separately. In a realistic case the trader would face both. We now assess the returns under conditions that an institutional investor might face when practically implementing a pairs trading strategy. We will specify a 60 second delay and deduct transaction costs of 3 bps. per trade. In addition we specify short fees equal to 500 bps. annually. The table on next page shows the results given these conditions.

The Top 5 pairs portfolio with high values specified for entry and stop-loss thresholds perform well under this setup, delivering high annualized returns with low annual standard deviations. The same is true for the portfolios comprised of the Top 10 pairs, given that the entry level is set to a level above two standard deviations. Sharp ratio values exceeding 3 indicate that the pairs trading stagey is able to generate abnormal risk adjusted returns.

The annualized returns to the OSEBX index over the sample period was 9.6 % with a corresponding standard deviation of 10.6 %. These figures results in a Sharpe ratio of 0.63. In a longer perspective, the average annual Sharpe ratio for the Norwegian stock market in the years 1900 – 2010 was 0.22. (Dimson, March and Staunton, 2011).

Table 6.3 – Controlling for Transaction Costs and Timing Constraints

Portfolio Parameters			Top 5 Pairs				Top 10 Pairs				All Pairs			
No	SL	q	Return	σ	SR	Trades	Return	σ	SR	Trades	Return	σ	SR	Trades
38	2 %	2	-10.02 %	6.0 %	-2.2	166	-11.83 %	5.2 %	-2.9	268	-8.85 %	3.8 %	-3.1	367
39	2 %	3	3.19 %	3.7 %	0.1	83	-2.15 %	3.0 %	-1.7	138	-2.48 %	2.4 %	-2.3	178
40	2 %	5	12.09 %	3.0 %	3.0	35	6.03 %	2.2 %	1.4	53	1.32 %	1.7 %	-1.0	62
42	5 %	2	1.28 %	4.7 %	-0.4	171	-2.12 %	4.6 %	-1.1	275	-2.87 %	3.6 %	-1.6	374
43	5 %	3	3.95 %	3.6 %	0.3	83	1.47 %	3.4 %	-0.5	138	0.30 %	2.6 %	-1.0	178
44	5 %	5	12.32 %	3.0 %	3.1	35	9.53 %	2.9 %	2.2	53	3.41 %	2.0 %	0.2	62
46	None	2	1.28 %	4.7 %	-0.4	171	-2.46 %	4.7 %	-1.2	275	-3.07 %	3.7 %	-1.7	374
47	None	3	3.95 %	3.6 %	0.3	83	1.11 %	3.4 %	-0.5	138	0.10 %	3.4 %	-0.8	178
48	None	5	12.32 %	3.0 %	3.1	35	9.53 %	2.9 %	2.2	53	3.41 %	2.0 %	0.2	62

Notes: This table presents the annualized returns on a pairs trading strategy after controlling for transaction costs and timing restrictions.

No Portfolio parameter identification number

SL Stop-loss threshold

q Threshold for opening a position. Numbers of standard deviations.

σ Annualized standard deviation on portfolio

SR Annualized Sharpe ratio.

Trades Number of roundtrips during the 3M testing period. One roundtrip translates to four trades. Two when opening a position and two when the position is closed.

6.6.4 Trade slippage

In addition to the direct trading costs we have examined, there exists another, more subtle, trading cost that is hard to control for. This cost is the result of the market impact caused by the trader when a position is opened or closed. This is commonly referred to as trade slippage and is defined by H. Zhang & Q. Zhang (2006, p. 1512) as “[...] the spread between the expected price and the price actually paid”. In our study this translates to the difference between the price where the algorithm signals entry or exit, and the price at which the trader is able to execute the order. We can illustrate this concept with a simple example. Assume that the algorithm monitors the current ask price of a stock and that a profitable opportunity at price 100 NOK is identified. A trade signal is issued instructing the trader to buy 100 shares. However, in the order book there might be only be 50 shares offered at 100 NOK. The remaining shares will have to be bought at the next best (or third best etc.) price available and therefore pushes up the average price per share.

The costs due to the slippage effect will be more of a problem in smaller less liquid stocks and when the positions traded are large. We argue that slippage would have

a low impact on the results presented in this study. The background for this is that our universe of stocks is limited to the 25 most liquid stocks listed on the Oslo Stock Exchange. Furthermore, the impact of slippage can be approximated in terms of higher transaction costs per trade. Our examination of transaction costs shows that even after deducting commissions of 10 basis points per trade, the strategy still results in annual returns just short of ten percent. This level of returns indicates that the returns to strategy are likely to be positive even after a moderate increase in costs due to slippage (or for that matter, any other cost not controlled for in this study). That being said, slippage would be a limiting factor when trying to implement the strategy at larger scales. The investor would eventually cause market impacts so large that all profits are eliminated.

6.8 Exposure to market risk

In previous studies (Gatev et al. 2006; Hoel 2013) it is documented that the exposure to market risk is close to zero for a pairs trading strategy. As mentioned in the introduction, the exposure to market risk depends on the amount of capital placed in each leg of a pair, and the market risk exposure in the stocks included in the position.

We analyze the exposure to market risk using the CAPM²⁸ framework. We compare the returns to the pairs trading strategy (before transaction costs) with the returns to the OSEBX²⁹ index. Our results are in line with the results previous literature. The results strongly suggest that the pairs trading strategy, as it is implemented in this paper, is insignificantly exposed to market risk. All the coefficient estimates are close to zero. In addition none of the coefficients are statistically significant even at a 20 % significance level. Given that the strategy involves simultaneously buying and selling stocks – from the same sector and in equal amounts – the results are not surprising. Intuitively if the two stocks share a similar exposure to market risk the exposure in the long position should be offset by the exposure in

²⁸ The Capital Asset Pricing Model is a model for determining the required rate of return for an asset. The model was introduced by Treynor, Sharpe, and Mossin in the mid-1960s.

²⁹ This index is a weighted average of all stocks listed on the Oslo Stock Exchange.

the short position. The risk exposure coefficients for the stocks considered in the restricted case are included in appendix J.

Table 6.4 – Exposure to market risk – Restricted Case

Portfolio Parameters			Top 5 Pairs				Top 10 Pairs				All Pairs			
No.	SL	q	Return	σ	β	t-value	Return	σ	β	t-value	Return	σ	β	t-value
38	2 %	2	5.49 %	5.7 %	-0.03	-0.36	-2.37 %	5.1 %	-0.03	-0.33	-0.59 %	3.7 %	-0.03	-0.52
39	2 %	3	4.53 %	3.5 %	0.03	0.43	-1.67 %	3.4 %	0.03	0.59	-1.00 %	2.4 %	0.01	0.27
40	2 %	5	16.53 %	2.8 %	-0.03	-0.54	8.81 %	2.0 %	0.00	0.10	3.40 %	1.4 %	-0.01	-0.49
42	5 %	2	16.95 %	4.6 %	-0.04	-0.51	7.86 %	4.7 %	-0.08	-0.97	5.58 %	3.5 %	-0.06	-1.09
43	5 %	3	9.40 %	3.6 %	0.01	0.16	5.74 %	3.4 %	-0.02	-0.34	4.03 %	2.5 %	-0.03	-0.64
44	5 %	5	16.87 %	2.9 %	-0.02	-0.44	12.61 %	2.8 %	-0.03	-0.72	5.66 %	1.8 %	-0.03	-1.08
46	None	2	16.95 %	4.6 %	-0.04	-0.51	7.71 %	4.8 %	-0.08	-0.98	5.50 %	3.5 %	-0.06	-1.09
47	None	3	9.40 %	3.6 %	0.01	0.16	5.60 %	4.4 %	-0.02	-0.36	3.95 %	2.5 %	-0.03	-0.65
48	None	5	16.87 %	2.9 %	-0.02	-0.44	12.61 %	2.8 %	-0.03	-0.72	5.66 %	1.8 %	-0.03	-1.08

Notes: This table presents the exposure to market risk for a pairs trading strategy.

No.	Portfolio parameter identification number
SL	Stop-loss Threshold
q	Threshold for opening a position. Numbers of standard deviations.
σ	Annualized standard deviation on portfolio
β	Coefficient of exposure to market risk according to the CAPM
t-value	Test statistic used to determine whether the coefficient values are statistically significant or not.

In addition to testing the exposure to market risk in the restricted case, we repeated the tests for the unrestricted case, where pairs can be formed across industries. The results from these tests are very similar to the ones discussed above: the coefficient values are small and not statistically significant³⁰.

It is important to note that even though exposure to market risk is negligible, there are several other sources of risk affecting a long short strategy such as this. As an example, imagine a scenario where a trader holds a long/short position and the company in the long leg of the position goes bankrupt. This would result in substantial losses. Similarly, losses would occur if the company sold short increases rapidly in price as a response to a tender offer. However, due to their infrequency it is difficult to quantify the risks associated with such events.

³⁰ These results are not included in this thesis but are available upon request.

6.9 Is pairs trading a masked mean reversion strategy?

In the 1990 paper “*Evidence of Predicable Behavior of Security Returns*” N. Jegadeesh show that following a strategy where past winners are sold short and past losers are bought long yields abnormal returns. Since then various forms of mean reversion strategies have been popular with investors. In the light of this it is natural to ask if pairs trading is merely an exotic variant of a mean reversion strategy. This question has been explored in previous literature. Gatev et al. (2006) argue that the excess results found in their study cannot be explained by mean reversion. This view is based on the results from two tests. The first approach is a factor analysis of the returns. The exposure to reversion and momentum factors are too small to fully explain the positive results. The second test involves applying a bootstrap technique. When the algorithm indicates that it is time to take a position in a pair both stocks are substituted for a pair where the stocks are randomly selected. Each stock is substituted with a random stock from the same return decile, measured over the previous month. This is analog to following a mean reversion strategy with the times for entry and exit determined by the pairs trading algorithm. When the bootstrapping procedure is implemented the returns are significantly lower than when trading the true pairs. Based on the results from the two outlined tests, the authors conclude that the mechanisms behind a pairs trading strategy and a pure mean reversion strategy are fundamentally different.

We support the view held by the authors in the discussed paper. In addition we would like to point out that pairs trading do not necessarily rely on mean reversion in the stock prices. It is the spread between the two stocks that need to exhibit mean reverting properties. A pairs trading strategy can yield positive results even if both stocks are trending strongly in the same direction, exhibiting properties normally associated with momentum in returns. The following figure illustrates such a scenario.

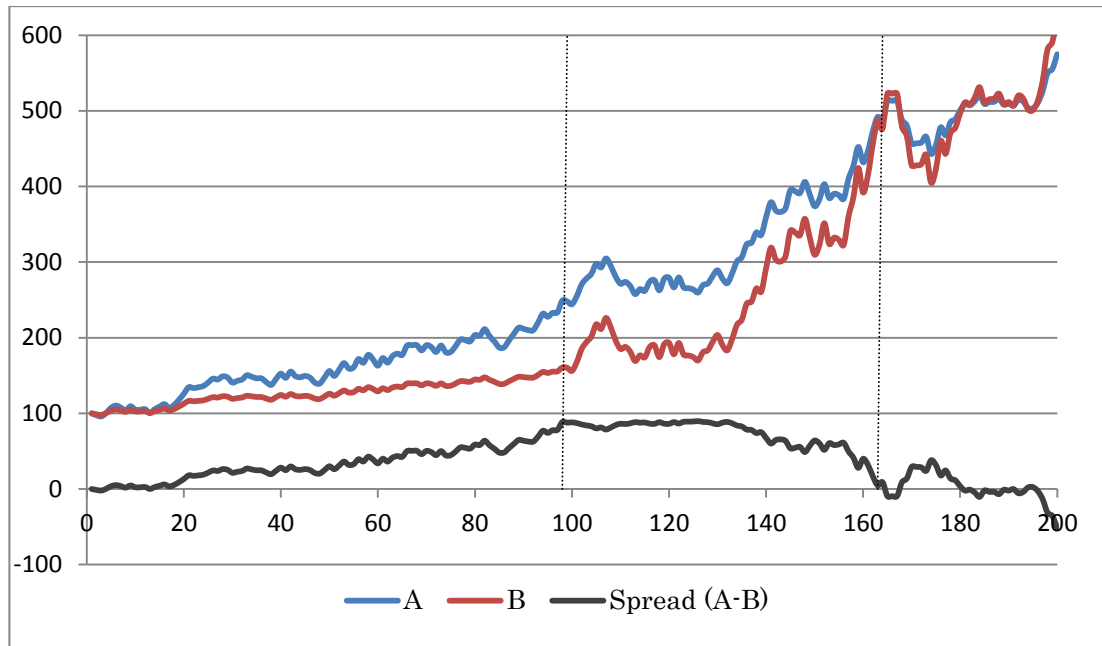


Figure 6.4 – A hypothetical pairs trading scenario with stocks exhibiting momentum

In this hypothetical situation the prices of stock A and stock B are both increasing. However, in the start of the series the price of stock A increases more rapidly than the price of Stock B. This leads to divergence and a position where A is shorted and B bought long is entered. The position is opened at the time marked by the first vertical line. After entering the position both stock prices climb at faster rates than before. However the price of B is increasing even faster than the price of A. After some time the price of stock B has increased to a point where the two stocks are again equal in value. The position is unwound and a profit is realized. The gains from the long position in stock B more than offset the losses from the short position in stock A. This illustrates that while mean reversion in stock prices could lead to profits when following a pairs trading strategy, it is not a necessary condition for the strategy to be successful.

7. Conclusion

In this paper we have explored various methods for implementing a pairs trading strategy. We conduct tests on simulated data in order to determine the optimal approach to pairs trading. In addition we conduct an empirical test of a pairs trading strategy in the Norwegian stock market using high frequency data. We find that a strategy where pairs are matched by considering past statistical relationships alone is unprofitable. Refining the formation procedure, by allowing only stocks from the same industry sector to be matched, we find that the strategy is able to generate significant returns with low standard deviations. This results in attractive Sharpe ratio values, indicating that the strategy is able to generate abnormal risk-adjusted returns. The results are robust to moderate transaction costs and the returns are uncorrelated with those of the general market, represented by the OSEBX index.

The positive returns obtained are generated by analyzing past price information. We also add a crude fundamental component to the strategy when we restrict the stocks in a pair to belong to the same sector. According to the Efficient Market Hypothesis (EMH) it should be impossible to obtain risk-adjusted returns better than the average market return. Our results could therefore indicate that the semi-strong form of the EMH did not hold in the Norwegian stock market over the period studied. Another explanation, as suggested by Gatev et al. (2006), could be that the returns are attributable to exposure to an unknown systematic risk factor. Furthermore our sample covers a relatively short period of time and we test a number of parameter configurations. It is therefore necessary to replicate the results using data from different time periods before concluding that the EMH is violated.

References

- Bossaerts, P., & Green, R. (1989). A General Equilibrium Model of Changing Risk Premia: Theory and Evidence. *Review of Financial Studies* 2, 467-493.
- Bowen, D., Hutchinson, M. C., & O'Sullivan, N. (2010). High-Frequency Equity Pairs Trading: Transaction Costs, Speed of Execution, and Patterns in Returns. *The Journal of Trading* 5, no. 3, 31-38.
- Caldeira, J. F., & Moura, G. V. (2013). Selection of a Portfolio of Pairs Based on Cointegration: A Statistical Arbitrage Strategy (Working Paper). Federal University of Rio Grande do Sul.
- Dimson, E., Marsh, P., & Staunton, M. (2011). Equity Premia Around the World. London Business School .
- Do, B., & Faff, R. (2010). Does Simple Pairs Trading Still Work? *Financial Analyst Journal* 66, no. 4, 83-95.
- Do, B., & Faff, R. (2012). Are Pairs Trading Profits Robust to Trading Costs. *Journal of Financial Research* 35, no. 2, 261-287.
- Do, B., Faff, R., & Hamza, K. (2006). A New Approach to Modeling and Estimation for Estimation for Pairs Trading. *Working Paper*. Monash University.
- Elliot, J. R., Hoek, J. v., & Malcolm, W. P. (2005). Pairs Trading. *Quantitative Finance* 5, no. 3, 271-276.
- Engelberg, J., Gao, P., & Jagannathan, R. (2009). "An Anatomy of Pairs Trading: The Role of Idiosyncratic News, Common Information and Liquidity." Working paper. University of California, University of Notre Dame, Northwestern University.
- Fama, E. F., & French, K. R. (1997). Industry costs of equity. *Journal of Financial Economics*, 153-193.
- Fama, E., & French, K. (1992). The Cross-section of Expected Stock Returns. *Journal of Finance*, 47, 427-486.
- Gatev, E., Goetzmann, W. M., & Rouwenhorst, K. G. (2006). Pairs Trading: performance of a Relative Value Arbitrage Rule. *The Review of Financial Studies* 19, no. 3, 797-827.
- Gregory, I., Knox, P., & Oliver-Ewald, C. (2011). Analytical Pairs Trading Under Different Assumptions on the Spread and Ratio Dynamics (Working paper). University of Sydney.
- Harris, R., & Sollis, R. (2003). *Applied Time Series Modelling and Forecasting*. Wiley.
- Hoel, C. H. (2013). Statistical Arbitrage Pairs: Can Cointegration Capture Market Neutral Profits? (Master's thesis). Bergen: Norwegian School of Economics.

- Jagannathan, R., & Viswanathan, S. (1988). Linear Factor Pricing, Term Structure of Interest Rates and the Small Firm Anomaly, Working Paper 57. Northwestern University.
- Jegadeesh, N. (1990). Evidence of Predicable Behavior of Security Returns. *The Journal of Finance*, Vol. 45 No 3., 881-898.
- Lamont, O. A., & Thaler, R. H. (2003). The Law of One Price in Financial Markets. *Journal of Economic Perspectives*, 191-202.
- Lin, Y.-X., McCrae, M., & Gulati, C. (2006). Loss Protection in Pairs Trading Through Minimum Profit Bounds: A Cointegration Approach. *Journal of Applied Mathematics and Decision Sciences*, 1-14.
- 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*; Vol. 6, No. 3, 96-110.
- Nath, P. (2003). High Frequency Pairs Trading with U.S Treasury Securities: Risks and Rewards for Hedge Funds (Working Paper). London Business School.
- Neumaier, A., & Schneider, T. (1997). Multivariate Autogressive and Ornstein—Uhlenbeck Processes: Estimates for Order, Parameters, Spectral Information, and Confidence Regions. *ACM Trans. Math. Software* 44, 83-155.
- Ross, S. (1976). The arbitrage theory of capital asset pricing. *Journal of Economic Theory* 13, 341–60.
- Steele, M. J. (2014, 6 10). *PairsTrading*. Retrieved from stat.wharton.upenn.edu/~steele:
<http://www-stat.wharton.upenn.edu/~steele/Courses/434/434Context/PairsTrading/PairsTrading.html>
- Vidyamurthy, G. (2004). *Pairs Trading: Quantitative Methods and Analysis*. New Jersey: John Wiley & Sons.
- Zhang, H., & Zang, Q. (2008). Trading a mean—reverting asset: Buy low and sell high. *Automatica* Vol. 44, 1511-1518.

Appendices

Appendix A – Logarithmic spread

An alternative approach to the normalization procedure is to compute the spread as the difference of the natural logarithms of the prices. This is possible because of the fact that the price at any point in time is equal to the price at t_0 , multiplied by the cumulative returns up until time t .

$$\begin{aligned} P_t &= P_0 * (1 + r_{t-1}) * (1 + r_{t-2}) * \dots * (1 + r_1) \\ P_t &= P_0 * r_t^{cum} \end{aligned} \tag{A1}$$

The product rule for logarithms states that $\ln(x * y) = \ln(x) + \ln(y)$. Therefore, assuming that $r_{At}^{cum} = r_{Bt}^{cum}$ we have that

$$\begin{aligned} \ln(P_{At}) - \ln(P_{Bt}) &= \ln(P_{A0} * r_{At}^{cum}) - \ln(P_{B0} * r_{Bt}^{cum}) \\ &= \ln(P_{A0}) + \ln(r_t^{cum}) - [\ln(P_{B0}) + \ln(r_t^{cum})] \\ &= \ln(P_{A0}) - \ln(P_{B0}) \end{aligned} \tag{A2}$$

We see that the spread of the logarithms is constant and equal to the spread at $t=0$, and that this holds true even for different levels of P_{A0} and P_{B0} . Thus the logarithmic transformation secures a consistent measure of the relative value development of the two stocks.

Appendix B – Estimating the parameters of the Ornstein–Uhlenbeck process using OLS³¹

Here we give an example of how to estimate the parameters of an Ornstein–Uhlenbeck process when we observe a series of realizations created by the process.

The stochastic differential equation used in this example is given by

$$dX_t = \lambda(\mu - X_t)dt + \sigma dW_t \quad \text{B1}$$

Where λ is the speed in which the series reverts to the long term mean μ . The volatility of the process is given by σ . W_t is a standard wiener process.

The solution to the stochastic differential equation given by the following

$$X_{i+1} = X_i e^{-\lambda\delta} + \mu(1 - e^{-\lambda\delta}) + \left[\sigma \sqrt{\frac{1 - e^{-2\lambda\delta}}{2\lambda}} \right] * n \quad \text{B2}$$

The function is continuous and δ is the fixed time step for each sample. n is normally distributed noise term with mean and standard deviation equal to zero and one respectively.

The observed values are listed in the table below. In addition we also list the specific random terms used to create the series. The following parameter values are used in this example. $\lambda = 1$, $\mu = 1$, $\sigma = 0.5$, $\delta = 1$.

³¹ Example adapted from <http://www.sitmo.com/article/calibrating-the-ornstein-uhlenbeck-model>

i	N~(0,1)	X(i)
0	-0.8541	3.0000
1	1.0821	2.7976
2	-1.4730	2.0733
3	1.7487	2.2237
4	0.7765	2.1252
5	-1.4525	1.5542
6	0.3774	1.5153
7	0.2297	1.4523
8	1.0682	1.5891
9	-0.7589	1.2905
10	0.4854	1.3339
11	-0.2990	1.1937
12	-1.1972	0.8854
13	-0.1042	0.8876
14	1.3718	1.2167
15	-1.7744	0.7753

We now want to estimate the parameters of the process given the observations we have of X. We estimate the following relationship with OLS. $X_{i+1} = \hat{\alpha} + \hat{\beta}X_i + \hat{\varepsilon}_{i+1}$.

The parameters are then found by analyzing the regression results:

$$\hat{\lambda} = -\frac{\ln(\hat{\beta})}{\hat{\delta}}, \quad \hat{\mu} = \frac{\hat{\alpha}}{1-\hat{\beta}} \text{ and } \hat{\sigma} = sd(\varepsilon) * \sqrt{\frac{-2\ln(\hat{\beta})}{\hat{\delta}(1-\hat{\beta}^2)}}.$$

Applying the above procedure on the dataset in this example gives us the following estimates.

Parameter	Estimate	True value
λ	0.9406	1
μ	0.9681	1
σ	0.5601	0.5

Appendix C – Conditional probability density function of the Ornstein–Uhlenbeck process ³²

The conditional probability distribution function for the process described in Appendix 2a is given by

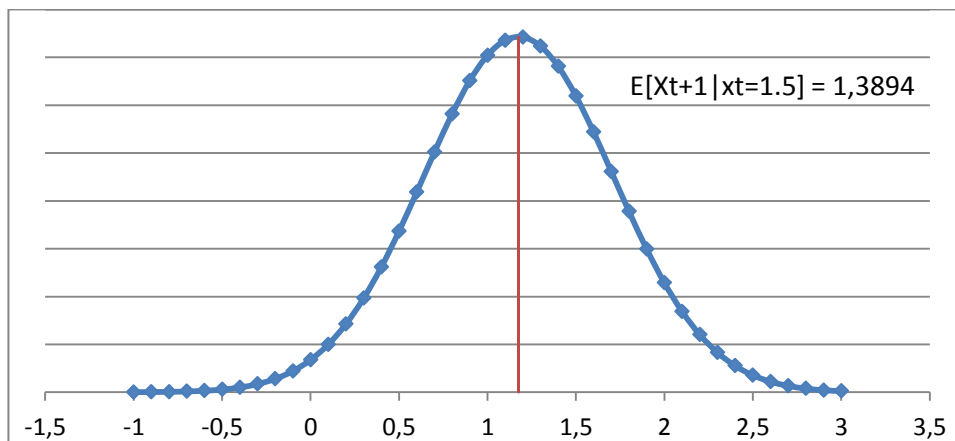
$$Cpdf = \frac{1}{\sqrt{2\pi\hat{\sigma}^2}} * \exp\left[-\frac{(X_i - X_{i-1}e^{-\lambda\delta} - \mu(1 - e^{-\lambda\delta}))^2}{2\hat{\sigma}^2}\right] \quad C1$$

Where $\hat{\sigma}^2 = \sigma^2 \frac{1 - e^{-2\lambda\delta}}{2\lambda}$.

The conditional expectation is found using the expression:

$$E[X_t | x_{t-1}] = \mu + (x_{t-1} - \mu) * e^{-\lambda\delta} \quad C2$$

In the figure below we apply C1 and C2 using the parameter values from appendix B. The previous observation X_{i-1} is set to 1.5 in this example.



³² Example adapted from <http://www.math.ku.dk/~susanne/StatDiff/Overheads1b>

Appendix D – Python code – Pairs identification procedures

1. Pairs ranking using the Distance approach

```
import statsmodels.api as sm
import statsmodels.tsa.stattools as ts
import math
import numpy as np
import scipy.odr.odrpack as odrpack
import scipy.stats
import itertools

def sumsq(f1,f2):
    """This function returns the sum of squared differences between two lists, in addition the
    standard deviation of the spread between the two lists are calculated and reported"""

    Y = f1 #Define each series
    X = f2
    spread = [] #Initialize variables
    std = 0
    cumdiff = 0

    for i in range(len(X)): #Calculate and store the sum of squares
        cumdiff += (X[i]-Y[i])**2
        spread.append(X[i]-Y[i])

    std = np.std(spread) #Calculate the standard deviation
    return(cumdiff,(std))
```

2. Pairs ranking using the Cointegration Coefficient approach - OLS regression

```
import statsmodels.api as sm
import statsmodels.tsa.stattools as ts
import math
import numpy as np
import scipy.odr.odrpack as odrpack
import scipy.stats
import itertools

def coint(f1,f2):
    """This function takes two lists and returns the DF-Test statistics"""

    Y = f1 #Define each series
    X = f2

    for i in range(len(X)): #Ln transform the prices
        X[i] = math.log(X[i])
        Y[i] = math.log(Y[i])

    X = sm.add_constant(X) #Set mode to regression with constant term

    model = sm.OLS(Y,X) #Specify the model to be used
    results = model.fit() #Run the regression

    R = results.resid #Store residuals
```

```

intercept = results.params[0]           #Store Intercept value
beta = results.params[1]               #Store coefficient value
rsq = results.rsquared                 #Store Rsquared

adf = ts.adfuller(R,0,"c",None,True,True) #Run dickey fuller test on the obtained
                                         residuals

z = adf[0]                             #Store test statistics
pval = adf[1]

return(z,pval,beta)

```

3. Pairs ranking using the Cointegration Coefficient approach - ODR regression

```

import statsmodels.api as sm
import statsmodels.tsa.stattools as ts
import math
import numpy as np
import scipy.odr.odrpack as odrpack
import scipy.stats
import itertools

def odrcoint(y,x):
    """This function takes two lists and returns the DF-Test statistics"""

    for i in range(len(x)):             #Ln transform the prices
        x[i] = math.log(x[i])
        y[i] = math.log(y[i])

        def f(B, x):                    #Define the model to be estimated
            return B[0] + B[1]*x

        linear = odrpack.Model(f)        #Regress y on x (Y = a +bx)
        mydata = odrpack.RealData(x, y, sx=1, sy=1)

        myodr = odrpack.ODR(mydata, linear, beta0=[0,-1])
        myoutput = myodr.run()

        intercept = myoutput.beta[0]    #Store the regression coefficients
        beta = myoutput.beta[1]

    resid = []

    for i in range(len(x)):             #Calculate the residuals
        est = intercept +(beta*x[i])
        res = y[i] - est
        resid.append(res)

    adf = ts.adfuller(resid,0,"c",None,True,True) #Run dickey fuller test on the obtained
                                                    residuals

    z = adf[0]                           #Save test statistics
    pval = adf[1]

    return(z,pval,beta)

```

Appendix E – Resulting ranking of simulated data generated by the Granger representation theorem

Z-values and P values refer to the dickey fuller test. The γ values are interpreted as the percentage return in Y given a one percentage return in X.

Sum of squares rank		Cointegration rank (Reg y on x)				Cointegration rank (Reg x on y)				Orthogonal regression			
Group/Pair	Sum	Pair	Z-val.	P-val.	Est Y	Pair	Z-val.	P-val.	Est Y	Pair	Z-val.	P-val.	Est Y
10 1	9.8k	10 9	-72.7	0	1.00	10 9	-72.7	0.00	0.99	10 9	-72.9	0.00	1.003
10 2	9.9k	10 4	-71.2	0	1.00	10 4	-71.2	0.00	1.00	10 4	-71.3	0.00	0.999
10 8	9.9k	10 6	-71.0	0	0.99	10 6	-71.0	0.00	0.99	10 6	-71.3	0.00	0.998
10 4	9.9k	10 1	-70.8	0	1.00	10 1	-70.8	0.00	1.00	10 1	-70.9	0.00	1.000
10 9	10.0k	10 5	-70.5	0	1.00	10 5	-70.5	0.00	1.00	10 5	-70.7	0.00	1.001
10 7	10.1k	10 2	-70.3	0	0.99	10 2	-70.3	0.00	0.99	10 2	-70.6	0.00	1.002
10 6	10.1k	10 7	-69.8	0	1.00	10 7	-69.8	0.00	1.00	10 1	-70.0	0.00	0.998
10 10	10.1k	10 10	-69.6	0	0.99	10 10	-69.6	0.00	0.99	10 7	-69.9	0.00	1.002
10 3	10.2k	10 3	-69.2	0	0.99	10 3	-69.2	0.00	1.00	10 3	-69.3	0.00	0.998
10 5	10.2k	10 8	-68.3	0	0.99	10 8	-68.3	0.00	0.99	10 8	-68.5	0.00	1.002
9 1	10.3k	9 9	-59.5	0	1.00	9 9	-59.5	0.00	0.99	9 9	-59.6	0.00	1.003
9 2	10.3k	9 4	-58.2	0	1.00	9 4	-58.2	0.00	1.00	9 4	-58.3	0.00	0.999
9 4	10.3k	9 5	-58.0	0	1.00	9 5	-58.0	0.00	1.00	9 5	-58.1	0.00	1.001
9 9	10.4k	9 6	-57.9	0	0.99	9 6	-57.9	0.00	0.99	9 6	-58.1	0.00	0.998
9 7	10.4k	9 1	-57.7	0	1.00	9 1	-57.7	0.00	1.00	9 1	-57.8	0.00	1.000
9 8	10.5k	9 7	-57.3	0	1.00	9 7	-57.3	0.00	1.00	9 2	-57.5	0.00	1.003
9 6	10.5k	9 2	-57.3	0	0.99	9 2	-57.3	0.00	0.99	9 7	-57.3	0.00	1.002
9 5	10.6k	9 10	-56.9	0	0.99	9 10	-56.9	0.00	0.99	9 10	-57.2	0.00	0.997
9 10	10.6k	9 3	-56.6	0	0.99	9 3	-56.6	0.00	1.00	9 3	-56.7	0.00	0.997
9 3	10.7k	9 8	-55.6	0	1.00	9 8	-55.6	0.00	0.99	9 8	-55.8	0.00	1.003
8 9	11.7k	8 9	-47.7	0	1.00	8 9	-47.7	0.00	0.99	8 9	-47.8	0.00	1.004
8 7	11.7k	8 5	-46.9	0	1.00	8 5	-46.9	0.00	0.99	8 5	-46.9	0.00	1.001
8 1	11.8k	8 4	-46.5	0	0.99	8 4	-46.5	0.00	1.00	8 4	-46.6	0.00	0.998
8 2	11.9k	8 6	-46.2	0	0.99	8 6	-46.2	0.00	0.99	8 6	-46.4	0.00	0.997
8 4	11.9k	8 1	-46.2	0	1.00	8 1	-46.2	0.00	1.00	8 1	-46.2	0.00	1.000
8 5	12.0k	8 7	-46.2	0	1.00	8 7	-46.2	0.00	1.00	8 7	-46.2	0.00	1.003
8 6	12.1k	8 2	-45.8	0	0.99	8 2	-45.8	0.00	0.99	8 2	-46.0	0.00	1.004
8 10	12.2k	8 10	-45.7	0	0.98	8 10	-45.7	0.00	0.99	8 10	-45.9	0.00	0.996
8 8	12.2k	8 3	-45.5	0	0.99	8 3	-45.5	0.00	1.00	8 3	-45.5	0.00	0.996
8 3	12.3k	8 8	-44.4	0	0.99	8 8	-44.4	0.00	0.99	8 8	-44.6	0.00	1.004
7 7	15.3k	7 9	-36.5	0	1.00	7 9	-36.5	0.00	0.99	7 9	-36.5	0.00	1.006
7 9	15.3k	7 5	-36.1	0	1.00	7 5	-36.1	0.00	0.99	7 5	-36.2	0.00	1.002
7 5	15.4k	7 4	-35.4	0	0.99	7 4	-35.4	0.00	1.00	7 4	-35.4	0.00	0.997
7 1	15.6k	7 7	-35.4	0	1.00	7 7	-35.4	0.00	0.99	7 7	-35.4	0.00	1.004
7 2	15.8k	7 1	-35.2	0	1.00	7 1	-35.2	0.00	1.00	7 6	-35.3	0.00	0.995
7 4	15.9k	7 6	-35.1	0	0.98	7 6	-35.1	0.00	0.99	7 1	-35.2	0.00	1.000
7 10	16.0k	7 10	-34.9	0	0.98	7 10	-34.9	0.00	0.99	7 10	-35.1	0.00	0.995
7 6	16.0k	7 2	-34.8	0	0.99	7 2	-34.8	0.00	0.98	7 2	-34.9	0.00	1.005
7 3	16.3k	7 3	-34.7	0	0.99	7 3	-34.7	0.00	1.00	7 3	-34.8	0.00	0.994
7 8	16.4k	7 8	-33.8	0	0.99	7 8	-33.8	0.00	0.98	7 8	-33.9	0.00	1.006
6 5	26.9k	6 5	-24.2	0	0.99	6 9	-24.2	0.00	0.98	6 5	-24.3	0.00	1.003
6 7	27.1k	6 9	-24.2	0	1.00	6 5	-24.2	0.00	0.99	6 9	-24.3	0.00	1.012
6 9	27.3k	6 1	-23.3	0	1.00	6 1	-23.3	0.00	1.00	6 10	-23.4	0.00	0.989
6 1	28.1k	6 7	-23.3	0	1.01	6 7	-23.3	0.00	0.99	6 1	-23.4	0.00	0.999
6 10	28.6k	6 4	-23.3	0	0.99	6 4	-23.3	0.00	0.99	6 7	-23.4	0.00	1.009
6 2	28.9k	6 10	-23.2	0	0.96	6 10	-23.2	0.00	0.98	6 4	-23.4	0.00	0.995
6 6	29.2k	6 6	-23.2	0	0.97	6 6	-23.2	0.00	0.99	6 6	-23.3	0.00	0.990
6 4	29.4k	6 3	-22.9	0	0.98	6 3	-22.9	0.00	1.00	6 3	-23.0	0.00	0.989
6 3	29.5k	6 2	-22.8	0	0.98	6 2	-22.8	0.00	0.97	6 2	-23.0	0.00	1.010
6 8	30.1k	6 8	-22.3	0	0.99	6 8	-22.3	0.00	0.97	6 8	-22.5	0.00	1.012

5 5	50.2k	5 5	-16.7	0	0.99	5 5	-16.7	0.00	0.97	5 5	-16.8	0.00	1.007
5 7	52.6k	5 9	-16.5	0	1.00	5 9	-16.5	0.00	0.96	5 9	-16.6	0.00	1.025
5 9	52.7k	5 1	-16.0	0	0.99	5 1	-16.0	0.00	0.99	5 1	-16.0	0.00	0.999
5 1	54.2k	5 6	-15.8	0	0.94	5 6	-15.8	0.00	0.98	5 6	-16.0	0.00	0.981
5 2	55.6k	5 7	-15.7	0	1.01	5 7	-15.7	0.00	0.98	5 10	-15.9	0.00	0.978
5 10	55.7k	5 4	-15.7	0	0.97	5 4	-15.7	0.00	0.99	5 4	-15.8	0.00	0.991
5 6	57.7k	5 10	-15.7	0	0.93	5 10	-15.7	0.00	0.97	5 7	-15.8	0.00	1.017
5 4	58.2k	5 2	-15.5	0	0.97	5 2	-15.5	0.00	0.93	5 2	-15.7	0.00	1.021
5 3	58.6k	5 3	-15.4	0	0.96	5 3	-15.4	0.00	1.00	5 3	-15.4	0.00	0.977
5 8	59.0k	5 8	-15.1	0	0.98	5 8	-15.1	0.00	0.94	5 8	-15.3	0.00	1.024
4 5	120.4k	4 9	-10.4	0	1.01	4 9	-10.4	0.00	0.90	4 9	-10.5	0.00	1.064
4 9	129.8k	4 5	-10.2	0	0.97	4 5	-10.2	0.00	0.94	4 5	-10.4	0.00	1.015
4 1	130.6k	4 1	-10.0	0	0.98	4 1	-10.0	0.00	0.99	4 6	-10.1	0.00	0.952
4 2	131.3k	4 6	-9.9	0	0.86	4 6	-9.8	0.00	0.95	4 1	-10.1	0.00	0.996
4 7	134.2k	4 2	-9.7	0	0.93	4 2	-9.7	0.00	0.85	4 2	-10.0	0.00	1.052
4 10	145.9k	4 7	-9.5	0	1.02	4 7	-9.5	0.00	0.94	4 10	-9.6	0.00	0.941
4 6	146.5k	4 4	-9.5	0	0.93	4 4	-9.4	0.00	0.97	4 4	-9.6	0.00	0.979
4 4	148.4k	4 10	-9.3	0	0.82	4 10	-9.3	0.00	0.91	4 7	-9.6	0.00	1.041
4 8	148.8k	4 8	-9.2	0	0.94	4 8	-9.2	0.00	0.85	4 8	-9.5	0.00	1.059
4 3	166.3k	4 3	-8.8	0	0.90	4 3	-8.8	0.00	1.00	4 3	-8.9	0.00	0.944
3 2	244.9k	3 9	-7.3	0	1.02	3 9	-7.4	0.00	0.82	3 9	-7.5	0.00	1.132
3 5	250.2k	3 1	-7.2	0	0.96	3 1	-7.2	0.00	0.98	3 6	-7.3	0.00	0.905
3 1	251.2k	3 6	-7.1	0	0.76	3 2	-7.0	0.00	0.75	3 2	-7.3	0.00	1.097
3 9	268.5k	3 2	-6.9	0	0.87	3 6	-7.0	0.00	0.90	3 1	-7.2	0.00	0.988
3 7	277.9k	3 5	-6.8	0	0.93	3 5	-6.9	0.00	0.88	3 5	-7.0	0.00	1.026
3 4	297.0k	3 4	-6.7	0	0.87	3 7	-6.6	0.00	0.89	3 4	-6.8	0.00	0.954
3 6	302.5k	3 7	-6.6	0	1.03	3 4	-6.6	0.00	0.95	3 10	-6.7	0.00	0.887
3 10	305.3k	3 10	-6.4	0	0.68	3 8	-6.3	0.00	0.73	3 7	-6.7	0.00	1.079
3 8	308.7k	3 8	-6.2	0	0.87	3 10	-6.3	0.00	0.82	3 8	-6.5	0.00	1.121
3 3	389.1k	3 3	-5.7	0	0.80	3 3	-5.6	0.00	0.97	3 3	-5.8	0.00	0.895
2 2	435.2k	2 9	-5.2	0	1.04	2 9	-5.3	0.00	0.68	2 2	-5.5	0.00	1.180
2 1	491.9k	2 1	-5.1	0	0.91	2 2	-5.3	0.00	0.61	2 9	-5.5	0.00	1.290
2 5	567.8k	2 2	-5.1	0	0.77	2 1	-5.1	0.00	0.97	2 6	-5.3	0.00	0.819
2 9	590.9k	2 6	-5.0	0	0.60	2 6	-4.8	0.00	0.79	2 1	-5.2	0.00	0.971
2 7	592.4k	2 4	-4.7	0	0.76	2 7	-4.7	0.00	0.81	2 10	-4.8	0.00	0.807
2 4	614.0k	2 7	-4.7	0	1.05	2 4	-4.6	0.00	0.91	2 4	-4.8	0.00	0.900
2 10	626.5k	2 10	-4.5	0	0.49	2 5	-4.4	0.00	0.74	2 7	-4.8	0.00	1.152
2 8	637.1k	2 5	-4.3	0	0.84	2 8	-4.3	0.00	0.53	2 5	-4.6	0.00	1.080
2 6	684.7k	2 8	-4.1	0.01	0.74	2 10	-4.2	0.00	0.63	2 8	-4.5	0.00	1.301
2 3	892.9k	2 3	-3.7	0.04	0.63	1 2	-3.9	0.00	0.21	1 2	-3.9	0.00	3.100
1 2	1400.2k	1 10	-3.0	0.07	-0.08	2 3	-3.5	0.01	0.86	2 3	-3.8	0.00	0.808
1 5	2758.1k	1 2	-2.7	0.08	0.53	1 8	-2.9	0.04	0.00	1 10	-3.0	0.04	-0.209
1 8	2974.9k	1 6	-2.7	0.10	0.02	1 9	-2.9	0.04	0.10	1 8	-3.0	0.04	-85.776
1 1	3030.4k	1 4	-2.6	0.23	0.24	1 5	-2.1	0.25	0.23	1 9	-2.9	0.04	8.691
1 7	3129.4k	1 1	-2.1	0.31	0.58	1 7	-2.0	0.27	0.51	1 6	-2.7	0.07	0.027
1 10	3431.9k	1 8	-1.9	0.36	-0.01	1 1	-2.0	0.28	0.86	1 4	-2.6	0.10	0.302
1 3	4141.2k	1 3	-1.8	0.41	0.18	1 4	-2.0	0.31	0.90	1 1	-2.3	0.18	0.761
1 4	4693.3k	1 7	-1.8	0.43	0.82	1 6	-1.9	0.33	0.03	1 5	-2.2	0.21	2.099
1 9	4973.5k	1 9	-1.7	0.68	0.65	1 10	-1.9	0.35	-0.13	1 7	-2.1	0.25	1.427
1 6	7295.7k	1 5	-1.2	0.01	0.36	1 3	-1.6	0.49	0.29	1 3	-1.9	0.33	0.403

Notes: This table presents the results when ranking data generated by the Granger representation theorem.

Est Y Actual cointegration coefficient obtained when regressing one series on the other.

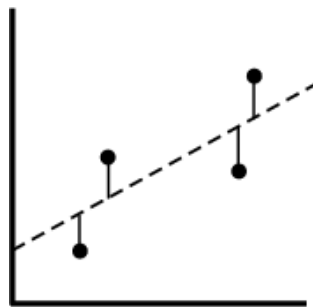
Sum Sum of squared differences between the series

Z-val Test statistic from the dickey fuller test

P-val P value for the test statistic observation

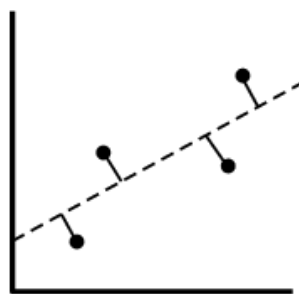
Appendix F – OLS vs orthogonal regression

The OLS algorithm minimizes the sum of squared vertical errors from the fitted line as illustrated in the figure below. The coefficients, test statistics and residuals are therefore sensitive to the ordering of the variables in the regression. This is a problem as the residuals are analyzed when determining whether two series are cointegrated or not.



OLS minimizes the vertical distance to the fitted line

Instead of minimizing the vertical distance to the fitted line the orthogonal regression minimizes the perpendicular distance. This is illustrated in the figure below. This approach yields coefficients, test statistics and residuals that are indifferent to the ordering of the variables.



Orthogonal regression minimizes the orthogonal distance to the fitted line

This type of regression is sometimes referred to as Total Least Squares or Deming regression in the two-variable case.

Appendix G –Ranking of simulated pairs generated with the common trends model

Z-values and P values refer to the dickey fuller test. The γ values are interpreted as the percentage return in Y given a one percentage return in X.

Sum of squares rank				Cointegration rank (Reg y on x)						Cointegration rank (Reg x on y)					
D_i	True γ	No.	Sum	D_i	True γ	No.	Z-Val	P-val	Est γ	D_i	True γ	No.	Z-Val	P-val	Est γ
0.1	1	1	5881k	0.1	0.25	2	-70.694	0	0.141	0.1	0.5	0	-72.953	0	0.517
0.1	1	2	6279k	0.1	0.1	0	-70.385	0	0.091	0.1	1	3	-71.691	0	1.002
0.1	1	0	6362k	0.1	0.25	4	-70.19	0	0.082	0.1	1	0	-71.678	0	0.938
0.1	1	3	6543k	0.1	0.1	4	-70.114	0	0.034	0.1	0.5	1	-71.525	0	0.46
0.5	1	1	6577k	0.1	0.25	0	-69.899	0	0.198	0.1	0.25	2	-71.349	0	0.247
0.1	0.5	1	6916k	0.1	0.1	2	-69.656	0	0.03	0.1	0.1	0	-70.989	0	0.201
0.1	0.5	3	7191k	0.1	0.1	1	-69.254	0	0.037	0.1	0.5	2	-70.603	0	0.538
0.1	1	4	7284k	0.1	0.5	0	-69.092	0	0.339	0.1	0.25	0	-70.499	0	0.231
0.1	0.5	2	7468k	0.1	0.5	1	-68.83	0	0.213	0.1	0.25	4	-70.464	0	0.146
0.1	0.1	3	7540k	0.1	0.25	3	-68.646	0	0.1	0.1	0.5	4	-70.342	0	0.41
0.5	0.5	1	7635k	0.1	0.5	4	-68.492	0	0.273	0.1	1	4	-70.114	0	1.334
0.5	0.5	3	7942k	0.1	0.25	1	-67.905	0	0.173	0.1	0.1	4	-70.111	0	0.099
0.5	1	2	7945k	0.5	0.25	0	-66.838	0	0.248	0.1	0.1	2	-69.731	0	0.071
0.5	0.5	2	8000k	0.1	0.1	3	-66.764	0	0.057	0.1	0.1	1	-69.336	0	0.067
1	1	1	8775k	0.1	0.5	2	-66.308	0	0.257	0.1	1	2	-69.331	0	0.934
0.5	0.25	2	9008k	0.5	0.25	2	-66.027	0	0.044	0.1	0.25	1	-69.27	0	0.297
0.5	0.5	0	9189k	0.1	1	1	-65.524	0	0.871	0.1	1	1	-69.136	0	0.986
1	0.5	1	9263k	0.1	0.5	3	-65.044	0	0.277	0.1	0.25	3	-69.114	0	0.16
0.5	0.1	2	9307k	0.5	0.5	1	-63.927	0	0.196	0.1	0.5	3	-68.252	0	0.554
0.1	0.1	4	9534k	0.5	0.1	2	-63.869	0	0.014	0.5	0.25	0	-67.603	0	0.287
0.1	0.5	0	9628k	0.5	0.1	4	-62.92	0	0.022	0.1	0.1	3	-67.329	0	0.189
1	0.1	2	9846k	0.5	0.5	0	-62.559	0	0.351	0.5	0.1	0	-66.216	0	0.545
1	0.1	0	11009k	0.5	0.1	0	-62.447	0	0.268	0.5	1	0	-66.135	0	0.766
0.5	0.1	0	11027k	0.1	1	3	-62.273	0	0.665	0.5	0.25	2	-65.935	0	0.105
0.5	1	3	11339k	0.1	1	0	-62.136	0	0.558	0.5	0.5	1	-65.89	0	0.431
0.1	0.1	2	11491k	0.5	0.5	2	-62.009	0	0.24	0.5	0.5	0	-65.761	0	0.521
1	0.5	0	11562k	0.5	0.1	1	-61.645	0	-0.062	0.5	0.5	2	-65.242	0	0.48
0.1	0.5	4	11637k	0.5	1	1	-60.773	0	0.834	0.5	1	4	-65.083	0	3.806
0.5	0.1	1	11663k	0.5	1	0	-60.329	0	0.53	0.5	0.1	2	-63.883	0	0.051
0.1	0.1	0	11694k	1	0.25	0	-59.736	0	0.3	0.5	1	1	-63.803	0	0.946
0.1	0.25	2	11748k	0.1	1	2	-59.699	0	0.5	0.5	1	3	-63.026	0	1.035
0.1	0.1	1	11868k	1	0.25	2	-58.877	0	-0.109	0.5	0.1	4	-62.91	0	0.035
0.5	1	0	12012k	0.5	1	2	-57.375	0	0.394	0.5	0.1	1	-61.865	0	-0.14
1	0.5	2	12500k	0.5	0.5	4	-57.184	0	-0.087	0.5	1	2	-61.864	0	0.655
1	0.5	3	12605k	0.5	0.25	1	-55.827	0	0.298	1	0.25	0	-60.557	0	0.347
0.5	0.1	3	12615k	0.5	1	3	-55.778	0	0.698	1	0.1	0	-60.19	0	0.87
1	1	2	13481k	0.1	1	4	-55.216	0	0.319	1	0.25	2	-59.976	0	-0.297
2	0.1	0	13695k	1	0.1	0	-53.742	0	0.491	0.5	0.25	1	-57.653	0	0.42
1	0.1	1	13798k	0.5	0.25	4	-53.144	0	-0.222	0.5	0.5	4	-57.426	0	-0.164
0.1	0.25	3	14244k	1	0.1	2	-52.799	0	-0.032	0.5	0.5	3	-56.776	0	0.875
0.1	0.25	1	14571k	1	0.5	2	-52.113	0	0.254	0.5	0.25	4	-55.037	0	-0.606
2	0.5	1	14951k	1	1	0	-52.038	0	0.468	1	1	0	-54.989	0	0.621
2	1	1	17602k	0.5	0.5	3	-52.028	0	0.255	0.5	0.1	3	-54.603	0	2.018
1	0.25	0	18569k	1	0.5	1	-51.816	0	0.23	1	0.5	2	-53.908	0	0.417
2	0.25	0	19985k	0.5	0.25	3	-50.938	0	-0.073	1	0.5	1	-53.895	0	0.53
3	0.1	0	20010k	1	1	1	-50.433	0	0.777	1	1	4	-53.366	0	106.283
0.5	0.25	0	21442k	1	0.5	0	-49.823	0	0.338	1	0.1	2	-53.093	0	-0.165
2	0.1	2	21816k	1	1	2	-49.794	0	0.242	1	1	1	-52.724	0	0.897
1	0.25	2	22329k	1	0.1	4	-48.561	0	0.033	1	0.5	0	-52.034	0	0.513
0.1	0.25	4	23868k	1	0.1	1	-47.874	0	-0.166	0.5	0.25	3	-51.075	0	-0.15

3	0.5	1	23881k	2	0.25	0	-44.038	0	0.37	1	1	2	-50.858	0	0.365
0.1	0.25	0	25460k	1	1	3	-43.363	0	0.688	1	0.1	1	-50.264	0	-0.594
1	1	3	25926k	2	0.25	2	-41.776	0	-0.215	1	0.1	4	-48.561	0	0.04
2	0.1	1	26043k	2	0.1	0	-41.24	0	0.779	1	1	3	-48.178	0	1.065
2	0.5	0	26048k	1	0.5	4	-41.157	0	-0.565	1	0.5	3	-47.765	0	5.247
0.5	0.1	4	26740k	0.5	0.1	3	-40.894	0	0.245	1	0.5	4	-46.557	0	-1.331
0.5	1	4	27415k	1	0.25	1	-40.438	0	0.376	2	0.1	0	-46.491	0	1.171
0.5	0.25	1	28714k	2	1	2	-38.98	0	0.042	2	0.25	0	-44.735	0	0.431
1	1	0	29613k	0.5	1	4	-38.936	0	0.237	2	0.5	3	-44.714	0	-4.029
3	0.25	0	30954k	1	0.5	3	-36.672	0	0.101	1	0.25	4	-44.711	0	-2.663
3	1	1	32336k	2	1	0	-36.348	0	0.358	1	0.1	3	-43.696	0	3.547
0.5	0.25	3	32649k	2	0.1	2	-35.56	0	-0.095	2	0.25	2	-42.493	0	-0.451
2	0.5	2	34282k	2	0.5	2	-35.082	0	0.283	1	0.25	1	-41.831	0	0.491
2	0.5	3	34344k	2	0.5	1	-34.256	0	0.352	2	1	2	-38.984	0	0.055
2	1	2	36056k	2	1	1	-33.804	0	0.653	2	0.25	4	-37.538	0	-13.797
0.5	0.5	4	37935k	1	0.25	4	-33.054	0	-0.511	2	1	0	-37.352	0	0.444
1	0.1	3	40666k	3	0.25	0	-32.726	0	0.407	2	0.5	4	-36.963	0	-3.896
0.5	0.25	4	46752k	3	1	2	-31.897	0	-0.055	2	0.5	1	-36.931	0	0.823
3	0.1	2	48311k	2	0.5	0	-31.72	0	0.277	2	0.1	2	-36.103	0	-0.329
3	0.1	1	48918k	3	0.1	0	-31.37	0	0.907	2	0.5	2	-35.786	0	0.378
3	0.5	0	53519k	1	0.25	3	-30.429	0	-0.167	3	0.5	3	-35.683	0	-2.9
1	0.25	1	57832k	2	0.1	4	-29.758	0	0.042	2	1	1	-35.279	0	0.803
1	0.1	4	72435k	3	0.25	2	-29.176	0	-0.202	2	0.1	1	-35.072	0	-2.125
1	0.25	3	72510k	2	0.1	1	-28.428	0	-0.203	3	0.1	0	-34.947	0	1.298
3	0.5	3	72633k	2	1	3	-26.954	0	0.616	3	0.25	0	-33.296	0	0.483
3	0.5	2	73104k	3	1	0	-26.922	0	0.28	2	0.5	0	-32.9	0	0.472
3	1	2	73973k	2	0.5	3	-26.526	0	-0.432	3	1	2	-31.923	0	-0.066
2	1	3	83740k	1	1	4	-26.363	0	-0.043	1	0.25	3	-31.805	0	-1.155
1	0.25	4	89110k	2	0.5	4	-25.8	0	-1.409	2	1	4	-31.75	0	-3.543
1	1	4	89573k	3	0.1	2	-25.746	0	-0.107	3	0.25	4	-31.676	0	98.286
2	1	0	99947k	1	0.1	3	-25.573	0	0.434	2	0.1	4	-29.76	0	0.047
1	0.5	4	104034k	3	0.5	2	-25.385	0	0.29	2	1	3	-29.608	0	1.119
2	0.25	2	104792k	3	0.5	1	-25.367	0	0.446	3	0.25	2	-29.318	0	-0.417
2	0.25	1	154201k	3	1	1	-24.626	0	0.544	3	0.5	4	-28.44	0	-7.447
2	0.1	3	169118k	2	0.25	1	-24.362	0	0.419	3	0.5	1	-27.824	0	0.983
3	1	3	179737k	3	0.5	3	-23.345	0	-0.726	3	1	0	-27.396	0	0.336
2	0.25	3	208417k	3	0.5	0	-22.581	0	0.219	3	0.1	2	-26.002	0	-0.285
3	1	0	217122k	3	0.1	4	-20.802	0	0.042	3	0.5	2	-25.788	0	0.361
2	0.25	4	219670k	3	0.1	1	-19.114	0	-0.118	3	1	1	-25.691	0	0.721
2	0.1	4	244451k	3	1	3	-19.074	0	0.551	2	0.1	3	-25.62	0	4.283
3	0.25	2	261680k	2	1	4	-18.037	0	-0.351	2	0.25	1	-25.23	0	0.521
3	0.25	1	301413k	3	0.25	1	-17.437	0	0.411	3	0.1	1	-25.068	0	-4.578
2	1	4	337197k	3	0.5	4	-16.944	0	-1.952	3	0.5	0	-23.3	0	0.416
2	0.5	4	346994k	2	0.25	4	-15.7	0	-0.551	2	0.25	3	-20.997	0	123.297
3	0.1	3	394038k	2	0.25	3	-15.031	0	-0.078	3	1	4	-20.99	0	-2.138
3	0.25	4	411354k	2	0.1	3	-13.971	0	0.492	3	1	3	-20.818	0	1.188
3	0.25	3	419236k	3	1	4	-13.392	0	-0.377	3	0.1	4	-20.805	0	0.046
3	0.1	4	523969k	3	0.25	3	-10.052	0	0.117	3	0.25	1	-18.064	0	0.495
3	0.5	4	737635k	3	0.25	4	-9.689	0	0.082	3	0.1	3	-17.284	0	4.163
3	1	4	749234k	3	0.1	3	-9.369	0	0.439	3	0.25	3	-16.068	0	27.526

Notes: This table presents the results when ranking data generated by the Stock & Watson common trends model.

D_i Sensitivity to specific non-stationary noise factor. Higher values results in series with random walk properties.

True γ The expected cointegration coefficient when regressing one series on the other.

Est γ Actual cointegration coefficient obtained when regressing one series on the other.

Appendix H – Adjustments of price series

Adjustments for dividends and splits		
<i>Ticker</i>	<i>Event date</i>	<i>Event</i>
GOGL	March 5 th 2014	Ex. Dividend 0.0025 USD
MHG	February 19 th 2014	Ex. Dividend 1.2 NOK
MHG	January 21 st 2014	Reverse split 10:1
PRS	February 14 th 2014	Ex. Dividend 0.16 USD
RCL	February 14 th 2014	Ex. Dividend 0.25 USD
SDRL	March 5 th 2014	Ex. Dividend 0.98 USD

Notes: This table presents the set of valid pairs under the restricted setup.

Appendix I – Python code implementing the pairs trading routine

```
def trade(AmidN,BmidN,Abidlist,Aasklist,Bbidlist,Basklist,threshold):
```

```
"""This function calculates the return generated from trading a pair of stocks over some time period. The function requires two lists of midprices for the two stocks as input. In addition four lists containing the actual bid/ask prices must be supplied. Lastly the entry threshold must be defined"""
```

```
    wait = 60                #Order execution delay parameter, seconds
    comission = 0.0003      #Comissionaction cost parameter
    shortcost = 0.05        #Short fee parameter

    shortlen = 0            #Initializing variables
    spread = 0
    prevspread = 0

    longpos = 0
    shortpos = 0
    trades = []

    longleg = ""

    openpos = False
    crossing = False
    Overval = ""
    slippage = 0

#Calculate the spread at each setof observatins
    for i in range(len(AmidN)-wait):
        spread = (AmidN[i]-BmidN[i])

#Trading signal for entering a trade
        if openpos == False:
            if abs(spread) > threshold:

#Heuristically control for comissionaction costs
            Transcost = ((Aasklist[i]-Abidlist[i])/2)+((Basklist[i]-Bbidlist[i])/2)+(4*comission)

                if Transcost < abs(spread):

                    if spread > 0:
                        Overval = "A"
                    else:
                        Overval = "B"

#Stock A is relatively overvaluedand therefore shorted. B bought long.
                    if Overval == "A":
                        longleg = "B"

#Must buy and sell stock at ask/bid price
                        short = Abidlist[i+wait]
                        long = Basklist[i+wait]

#Opposite of above case.
                    else:
```

```

        longleg = "A"

        short = Bbidlist[i+wait]
        long = Aasklist[i+wait]

    openpos = True
    #Check if exit conditions are fulfilled

    if openpos == True:

    #Calculate current position value
        if longleg == "B":
            longpos = 100*(Bbidlist[i]/long)
            shortpos = -100*(Aasklist[i]/short)
        if longleg == "A":
            longpos = 100*(Abidlist[i]/long)
            shortpos = -100*(Basklist[i]/short)

    #Stoploss function
        if longpos+shortpos<-10:

    #Calculate value of position considering bid/ask prices
        if longleg == "B":
    considering the bid/ask spread
            longpos = 100*(Bbidlist[i]/long)
            shortpos = -100*(Aasklist[i]/short)
        if longleg == "A":
            longpos = 100*(Abidlist[i+wait]/long)
            shortpos = -100*(Basklist[i+wait]/short)

    #Calculate short fee
        if i > 26400:
            shortlen = 2
        else:
            shortlen = 1

        trades.append((longpos*(1-comission))+(shortpos*(1+comission))-
            (2*comission*100)-((shortlen/365)*100*shortcost))

    return trades

    #Determine if prices cross at current set of observations
        if spread < 0 and prevspread > 0:

            crossing = True
        if spread < 0 and prevspread > 0:
            crossing = True

    #Exit position if prices cross
        if crossing == True:
            if longleg == "B":

    #Calculate value of position considering the bid/ask spread
            longpos = 100*(Bbidlist[i+wait]/long)
            shortpos = -100*(Aasklist[i+wait]/short)

            if longleg == "A":

```

```

        longpos = 100*(Abidlist[i+wait]/long)
        shortpos = -100*(Basklist[i+wait]/short)
#Calculate short fee
        if i > 26400:
            shortlen = 2
        else:
            shortlen = 1
#Store the result of the trade

        trades.append((longpos*(1-comission))+(shortpos*(1+comission))-
                      (2*comission*100)-((shortlen/365)*100*shortcost))

        openpos = False
        crossing = False

    prevspread = spread

#This is the end of the trading period. A position still open is automatically closed.

    if openpos == True:

#Calculate value of position considering the bid/ask spread

        if longleg == "B":
            longpos = 100*(Bbidlist[i+wait]/long)
            shortpos = -100*(Aasklist[i+wait]/short)
#Store the result of the trade
            trades.append((longpos*(1-comission))+(shortpos*(1+comission))
                          -(2*comission*100)-((2/365)*100*shortcost))
            openpos = False
        if longleg == "A":
#Store the result of the trade
            longpos = 100*(Abidlist[i+wait]/long)
            shortpos = -100*(Basklist[i+wait]/short)
            trades.append((longpos*(1-comission))+(shortpos*(1+comission))
                          -(2*comission*100)-((2/365)*100*shortcost))
            openpos = False

    return trades

```

Appendix J – Possible pairs in the restricted setup

Possible pairs in the Restricted Case				
<i>Stock A</i>	<i>Stock B</i>	<i>CAPM β Stock A</i>	<i>CAPM β Stock B</i>	<i>β difference (Absolute value)</i>
AKSO	PRS	1.57	1.2	0.37
AKSO	PGS	1.57	1.71	0.14
AKSO	SUBC	1.57	1.43	0.14
AKSO	TGS	1.57	1.49	0.08
DETN	PRS	1.21	1.2	0.01
DETN	PGS	1.21	1.71	0.5
DETN	SUBC	1.21	1.43	0.22
DETN	TGS	1.21	1.49	0.28
FOE	SDRL	1.19	0.99	0.2
MHG	ORK	1.15	0.73	0.42
PGS	SUBC	1.71	1.43	0.28
PGS	PRS	1.71	1.2	0.51
PGS	TGS	1.71	1.49	0.22
PRS	SUBC	1.2	1.43	0.23
PRS	TGS	1.2	1.49	0.29

Notes: This table presents the set of valid pairs under the restricted setup.

Stock A First stock in pair

Stock B Second stock in pair

CAPM β The associated market coefficient of the stock when assessed with the CAPM.

β difference The absolute difference in the CAPM β values when comparing stock A and stock B

Appendix K –Results in the restricted case controlling for transaction costs

Results Restricted Case – Impact of Transaction costs															
Portfolio Parameters				Top 5 Pairs				Top 10 Pairs				All Pairs			
No	SL	q	Fee	Return	σ	SR	Trades	Return	σ	SR	Trades	Return	σ	SR	Trades
38	2 %	2	0	5.49 %	5.7 %	0.4	170	-2.37 %	5.1 %	-1.0	272	-0.59 %	3.7 %	-1.0	371
39	2 %	3	0	4.53 %	3.5 %	0.4	83	-1.67 %	3.4 %	-1.4	138	-1.00 %	2.4 %	-1.7	178
40	2 %	5	0	16.53 %	2.8 %	4.8	35	8.81 %	2.0 %	2.9	53	3.40 %	1.4 %	0.3	62
42	5 %	2	0	16.95 %	4.6 %	3.0	171	7.86 %	4.7 %	1.0	275	5.58 %	3.5 %	0.7	374
43	5 %	3	0	9.40 %	3.6 %	1.8	83	5.74 %	3.4 %	0.8	138	4.03 %	2.5 %	0.4	178
44	5 %	5	0	16.87 %	2.9 %	4.8	35	12.61 %	2.8 %	3.5	53	5.66 %	1.8 %	1.5	62
46	None	2	0	16.95 %	4.6 %	3.0	171	7.71 %	4.8 %	1.0	275	5.50 %	3.5 %	0.7	374
47	None	3	0	9.40 %	3.6 %	1.8	83	5.60 %	4.4 %	0.6	138	3.95 %	2.5 %	0.4	178
48	None	5	0	16.87 %	2.9 %	4.8	35	12.61 %	2.8 %	3.5	53	5.66 %	1.8 %	1.5	62
38	2 %	1	3	-4.84 %	5.7 %	-1.4	170	-10.14 %	5.1 %	-2.6	272	-7.00 %	3.7 %	-2.7	371
39	2 %	2	3	-0.65 %	3.5 %	-1.0	83	-5.77 %	3.5 %	-2.5	138	-4.15 %	2.4 %	-2.9	178
40	2 %	3	3	14.04 %	2.8 %	4.0	35	7.08 %	1.9 %	2.1	53	2.26 %	1.4 %	-0.5	62
42	5 %	1	3	5.44 %	4.6 %	0.5	171	-0.81 %	4.7 %	-0.8	275	-1.27 %	3.5 %	-1.2	374
43	5 %	2	3	3.96 %	3.5 %	0.3	83	1.34 %	3.4 %	-0.5	138	0.74 %	2.5 %	-0.9	178
44	5 %	3	3	14.38 %	2.8 %	4.1	35	10.79 %	2.7 %	2.9	53	4.48 %	1.8 %	0.8	62
46	None	2	3	5.44 %	4.6 %	0.5	171	-0.95 %	4.7 %	-0.8	275	-1.35 %	3.5 %	-1.2	374
47	None	3	3	3.96 %	3.5 %	0.3	83	1.21 %	3.4 %	-0.5	138	0.66 %	2.5 %	-0.9	178
48	None	5	3	14.38 %	2.8 %	4.1	35	10.79 %	2.7 %	2.9	53	4.48 %	1.8 %	0.8	62
38	2 %	2	10	-21.85 %	5.6 %	-4.4	169	-23.28 %	5.1 %	-5.15	271	-18.07 %	3.7 %	-5.7	370
39	2 %	3	10	-9.92 %	3.5 %	-3.7	83	-9.92 %	3.5 %	-3.70	138	-9.89 %	2.5 %	-5.1	178
40	2 %	5	10	9.45 %	2.7 %	2.4	35	3.80 %	1.8 %	0.44	53	0.09 %	1.4 %	-2.0	62
42	5 %	2	10	-13.50 %	4.6 %	-3.6	170	-15.46 %	4.7 %	-3.96	274	-13.12 %	3.5 %	-4.6	373
43	5 %	3	10	-5.73 %	3.5 %	-2.5	83	-6.57 %	3.4 %	-2.83	138	-5.29 %	2.6 %	-3.2	178
44	5 %	5	10	9.78 %	2.7 %	2.5	35	7.40 %	2.6 %	1.72	53	2.27 %	1.8 %	-0.4	62
46	None	2	10	-13.50 %	4.6 %	-3.6	170	-15.58 %	4.7 %	-3.96	274	-13.19 %	3.5 %	-4.6	373
47	None	3	10	-5.73 %	3.5 %	-2.5	83	-6.70 %	3.4 %	-2.84	138	-5.36 %	2.6 %	-3.2	178
48	None	5	10	9.78 %	2.7 %	2.5	35	7.40 %	2.6 %	1.72	53	2.27 %	1.8 %	-0.4	62

Notes: This table presents the annualized returns on a pairs trading strategy after controlling for transaction costs.

No Portfolio identification number

SL Stop-loss threshold

q Threshold for opening a position. Numbers of standard deviations.

Fee Transaction costs for one single trade, number of basis points

Time Length of formation period : Length of trading period

σ Annualized standard deviation on portfolio

SR Annualized Sharpe ratio.

Trades Number of roundtrips during the 3M testing period. One roundtrip translates to four trades. Two when opening a position and two when the position is closed.

Appendix L –Results in the restricted case controlling for execution speed

Results Restricted Case – Impact of Order Execution Speed

Portfolio Parameters				Top 5 Pairs				Top 10 Pairs				All Pairs			
No	SL	q	Del.	Return	σ	SR	Trades	Return	σ	SR	Trades	Return	σ	SR	Trades
38	2 %	2	0	5.49 %	5.7 %	0.4	170	-2.37 %	5.1 %	-1.0	272	-0.59 %	3.7 %	-1.0	371
39	2 %	3	0	4.53 %	3.5 %	0.4	83	-1.67 %	3.4 %	-1.4	138	-1.00 %	2.4 %	-1.7	178
40	2 %	5	0	16.53 %	2.8 %	4.8	35	8.81 %	2.0 %	2.9	53	3.40 %	1.4 %	0.3	62
42	5 %	2	0	16.95 %	4.6 %	3.0	171	7.86 %	4.7 %	1.0	275	5.58 %	3.5 %	0.7	374
43	5 %	3	0	9.40 %	3.6 %	1.8	83	5.74 %	3.4 %	0.8	138	4.03 %	2.5 %	0.4	178
44	5 %	5	0	16.87 %	2.9 %	4.8	35	12.61 %	2.8 %	3.5	53	5.66 %	1.8 %	1.5	62
46	None	2	0	16.95 %	4.6 %	3.0	171	7.71 %	4.8 %	1.0	275	5.50 %	3.5 %	0.7	374
47	None	3	0	9.40 %	3.6 %	1.8	83	5.60 %	4.4 %	0.6	138	3.95 %	2.5 %	0.4	178
48	None	5	0	16.87 %	2.9 %	4.8	35	12.61 %	2.8 %	3.5	53	5.66 %	1.8 %	1.5	62
38	2 %	2	1	0.32 %	6.2 %	-0.4	166	-3.72 %	5.3 %	-1.3	268	-2.06 %	3.9 %	-1.30	367
39	2 %	3	1	9.35 %	3.8 %	1.7	83	2.71 %	3.1 %	-0.1	138	1.17 %	2.4 %	-0.78	178
40	2 %	5	1	14.86 %	2.2 %	5.3	53	7.98 %	2.2 %	2.2	53	2.61 %	1.7 %	-0.23	62
42	5 %	2	1	14.04 %	4.7 %	2.3	171	7.71 %	4.6 %	1.0	275	4.86 %	3.6 %	0.51	374
43	5 %	3	1	10.16 %	3.7 %	1.9	83	6.48 %	3.4 %	1.0	138	4.04 %	2.6 %	0.40	178
44	5 %	5	1	15.09 %	3.1 %	4.0	35	11.57 %	3.0 %	2.9	53	4.74 %	2.1 %	0.84	62
46	None	2	1	14.04 %	4.7 %	2.3	171	7.33 %	4.7 %	0.9	275	4.64 %	3.7 %	0.45	374
47	None	3	1	10.16 %	3.5 %	2.1	83	6.11 %	3.5 %	0.9	138	3.89 %	2.7 %	0.31	178
48	None	5	1	15.09 %	3.1 %	4.0	35	11.57 %	3.0 %	2.9	53	4.74 %	2.1 %	0.84	62
38	2 %	2	10	1.40 %	5.6 %	-0.3	167	-3.26 %	4.7 %	-1.3	269	-2.41 %	3.3 %	-1.65	365
39	2 %	3	10	8.53 %	3.6 %	1.5	82	2.82 %	2.9 %	-0.1	137	1.32 %	2.0 %	-0.84	176
40	2 %	5	10	8.86 %	2.2 %	2.6	35	7.24 %	2.6 %	1.6	53	3.02 %	1.7 %	0.01	62
42	5 %	2	10	10.33 %	4.7 %	1.6	169	4.61 %	4.2 %	0.4	273	1.03 %	3.3 %	-0.60	370
43	5 %	3	10	10.24 %	3.6 %	2.0	82	5.94 %	3.4 %	0.9	137	2.39 %	2.7 %	-0.23	177
44	5 %	5	10	8.98 %	2.3 %	2.6	35	7.76 %	1.9 %	2.5	53	2.44 %	1.9 %	-0.29	62
46	None	2	10	10.33 %	4.7 %	1.6	169	4.15 %	4.3 %	0.3	273	0.77 %	3.4 %	-0.66	370
47	None	3	10	10.24 %	3.6 %	2.0	82	5.47 %	3.5 %	0.7	137	2.12 %	2.7 %	-0.32	177
48	None	5	10	8.98 %	2.7 %	2.3	35	7.76 %	2.7 %	1.8	53	2.44 %	1.9 %	-0.29	62
38	2 %	2	60	3.08 %	4.4 %	0.0	162	1.49 %	4.2 %	-0.4	264	-0.84 %	3.1 %	-1.23	353
39	2 %	3	60	1.21 %	3.0 %	-0.6	75	0.39 %	3.2 %	-0.8	128	-1.41 %	2.5 %	-1.78	166
40	2 %	5	60	8.06 %	2.7 %	1.9	32	7.31 %	2.9 %	1.5	49	2.40 %	2.0 %	-0.30	57
42	5 %	2	60	3.95 %	4.2 %	0.2	164	2.45 %	4.4 %	-0.1	266	0.09 %	3.3 %	-0.88	335
43	5 %	3	60	2.38 %	2.8 %	-0.2	78	1.05 %	3.5 %	-0.6	131	-0.63 %	2.7 %	-1.34	169
44	5 %	5	60	8.17 %	2.6 %	2.0	32	6.93 %	2.9 %	1.3	49	2.19 %	2.1 %	-0.39	57
46	None	2	60	3.95 %	4.2 %	0.2	164	1.87 %	4.5 %	-0.2	266	-0.24 %	3.4 %	-0.94	355
47	None	3	60	2.38 %	2.8 %	-0.2	78	0.48 %	3.7 %	-0.7	131	-0.96 %	2.8 %	-1.40	169
48	None	5	60	8.17 %	2.6 %	2.0	32	6.93 %	2.9 %	1.3	49	2.19 %	2.1 %	-0.39	57

Notes: This table presents the annualized returns on a pairs trading strategy after controlling for delays in the speed of transaction.

No Portfolio identification number

SL Stop-loss threshold

q Threshold for opening a position. Numbers of standard deviations.

Del. The delay between trade signal and trade execution. Minutes

σ Annualized standard deviation on portfolio

SR Annualized Sharpe ratio.

Trades Number of roundtrips during the 3M testing period. One roundtrip translates to four trades