ISSN 0280-5316
ISRN LUTFD2/TFRT--5857--SE

Optimal Pairs Trading using Stochastic Control Approach
- A Critical Evaluation

Roman Galeev
Wang Ching Ho

Lund University
Department of Automatic Control
June 2010

Lund University
Department of Automatic Control
Box 118
SE-221 00 Lund Sweden

Document name
MASTER THESIS

Date of issue
June 2010

Document Number
ISRN LUTFD2/TFRT--5857--SE

Author(s)
Roman Galeev and Wang Ching Ho

Supervisor
Karl Larsson Department of Economics, Lund
Erik Johannesson Automatic Control, Lund
Charlotta Johnssson Automatic Control, Lund (Examiner)

Sponsoring organization

Title and subtitle
Optimal Pairs Trading using Stochastic Control Approapach (Optimering genom stokastisk reglering – en kritisk granskning)

Abstract
As of date, quantitative trading methods are a strong growing niche of trading. More and more sophisticated models are employed to chase investment opportunities. The question is whether this is over-engineering and if it adds value to practitioners. We have found an interesting quantitative pairs trading strategy that belong to the family of relative value strategies. The studied stochastic control approach has many shortcomings. Among those is pair selection method which is not defined at all. Furthermore it has not yet been applied to real market data which makes it interesting to see how well it performs against a more basic pairs trading strategy.
The purpose of this thesis is to define a suitable pair selection method that supports the theoretical framework of stochastic control and to test this selection strategy against a more basic pairs trading strategy on historical market data. Through communication with hedge fund officials insight was gained on how investment strategies are applied in the market. Secondary data was gathered through the Datastream database, and the different simulations were run in MATLAB. Mathematical and economical theories were gathered through various textbooks and articles, and the direction of the study was discussed and decided with the advice of the supervisors.
The method of stationarity through ADF test was found to be the best of the selection methods tested with the stochastic control approach. After finding proper time frames the stochastic control approach was benchmarked against a basic control strategy. The outcome shows that the specific quantitative strategy chosen for this study is flawed and therefore might not perform as well as it should with less assumptions made in the modeling. This is a sign of a possible over-engineering phenomenon that exists in the market in competition for investment opportunities.

Keywords
Pairs Trading, Market Neutral Strategy, Stochastic Control, Cointegration, Stationarity, Hedge Fund Investment Strategies.

Classification system and/or index terms (if any)

Supplementary bibliographical information

ISSN and key title
0280-5316

ISBN

Language
English

Number of pages
77

Recipient's notes

Security classification

http://www.control.lth.se/publications/

Acknowledgements

We would like to thank our supervisors Erik Johannesson and Karl Larsson for guiding us through this project. We would also like to thank Hossein Ashgharian and Charlotta Johnsson for creative ideas during the introductory phases of this work, and Supakorn Mudchanatongsuk and James Primbs for answering our questions regarding their work, enabling us to evaluate their strategy. Finally, we would like to thank our hedge fund officials, who wish to remain anonymous, for sharing their insights in quantitative investment strategies.

List of notations

Economic

AMF	Autorité des marchés financiers.
ASIC	Australian Securities & Investments Commission.
Arbitrage	Investment opportunities with low or no risk.
Backtesting	Test a strategy on historical data.
BaFin	Bundesanstalt für Finanzdienstleistungsaufsicht.
Convertible	A bond that can be converted into shares.
Fail to deliver	Connected to short selling. If the short seller cannot find a borrower of the shorted stock within the regulated time limit the trade is registered as fail to deliver.
FSA	Financial Services Authority.
Hedging	The practice of lower the investor's exposure against a certain asset by taking the opposite position in a similar asset.
High frequency trading	Automated electronic trading with reaction time measured in milliseconds.
Leverage	Commonly refers to the use of debt to increase exposure of investment which means increased potential return or loss. Derivatives can be used to achieve leverage without the use of debt.
Long	Being long on a stock means that the investor owns the stock.
Margin	Deposit acting as collateral to cover the broker's credit risks.
Market maker	The middle man between a buyer and seller in the market.
Option	A contract that gives the buyer the right to sell or buy an asset at a pre-determined price.
SEC	Securities and Exchange Commission.
Short	To be short on a stock means that the investor has sold borrowed stock that has to be returned when the trade is unwinded.
Spread	Usually refers to price difference.
Stop-loss	A trigger used in trading to limit losses when a position goes towards the wrong direction. When triggered the position is unwinded to prevent further losses.
Utility function	A function measuring the investor's satisfaction.

Mathematical

Stochastic process	A process where many different outcomes are possible, with different probabilities.
Stochastic control	A method of controlling and optimizing a stochastic process.
Markov property	The future of a variable with the Markov property is only dependent on its present state and not its past.
Wiener process	A process that a variable with the Markov property follows.
Ornstein-Uhlenbeck	One type of process that fluctuate around a fixed mean value.
Cointegration	A relationship between two time series.
Lognormal series	A series that becomes normally distributed after taking the logarithm of each element.
Brownian motion	Movement in a random way.
Supremum	The smallest number that is greater than every other one in a series.
Mean reverting process	A process that at some point will revert to its mean value.
Spurious regression	Implying that two variables are connected in some way, when they actually are not.
Quantile	Points/levels in a probability distribution.
Covariance	A measurement of how much two variables change together.

Contents

1 Introduction

1.1 Background

New developments within technology have automated more and more of what was once done by manual labour. Computers nowadays administrate tasks that require precision, quantitative capability and speed exceeding that of the human. This is clearly visible in the financial markets. As computer speed is increasing so is the demand for more accurate and sophisticated models in trading using quantitative strategies. The rise of electronic trading has made it even more attractive to use quantitative strategies. Many stock exchanges like New York Stock Exchange (NYSE), Euronext and London Stock Exchange (LSE) now offer electronic trading platforms to provide faster executions.[1][2]

Pairs trading is a market neutral trading strategy formally originating from the 1980s at Morgan Stanley.[3] The market neutrality implies that the strategy could be used at any time point in an economic cycle. Pairs trading can therefore be profitable regardless the direction of the market.

Pairs trading depends on two stocks following a similar pattern which means that the spread should be predictable, usually not for a period longer than two years.[4] The basic idea behind the strategy is to take advantage of the spread pattern. The spread is often assumed to follow a mean reverting pattern. This means that any drift from the long-term equilibrium will be corrected sooner or later unless the relationship between the two stocks has changed and a new spread equilibrium has established.

A drift from long-term equilibrium of the spread means that the opportunity of going long on the undervalued and short on the overvalued stock arises. This position is unwinded when the spread goes back towards its long-term equilibrium to take profit. The larger the divergence from long-term equilibrium is when entering a position the more likely it is to be profitable within a certain time frame. At the same time not as many trades will be made. The risk of the strategy lies in a shift of long-term equilibrium of the spread which will cause the position to be open for a longer time period and potentially initiating stop-losses.

There are many topics that must be covered before a successful implementation of this strategy can be made including:

[1] London Stock Exchange website
[2] New York Stock Exchange & Euronext website
[3] Vidyamurthy, 2004, pp. 73-74
[4] Wiktorsson, M., Associate Professor, mathematical statistics, Lund University (2010)

- How should a stock pair be chosen?
- How is divergence measured?
- How can we tell that the spread has converged?
- How do we know how much to buy and sell of each stock?
- How should we minimize the overall risk for the strategy?
- Should a stop-loss be implemented and at what level?

The most basic form of pairs trading trigger trades through threshold comparison. When a certain pre-determined threshold level has been reached, a position is opened. The level is decided through monitoring the evolvement of the spread for a given time period up to the current date. Multiples of standard deviation from the mean of the spread is generally used. When a threshold is triggered a long and short position is entered. The opportunity of higher return with a higher multiple is balanced by fewer opened position caused by the high multiple. Similar rules are made for closing the position. When the spread has reverted back to a certain threshold the position is unwinded.

There is the possibility for the spread to drift away. The use of stop-losses can therefore limit the losses if the long-term equilibrium has changed or the fundamental relationship between the stocks has broken.

1.2 Problem discussion

Pairs trading once started out as a simple strategy with simple rules. New and more sophisticated approaches have since then evolved out of the basic strategy to chase unexploited opportunities that are not covered by the more basic models. Naturally the question of over-engineering arises as a result of the previous statement. One danger that lies in creating complex models is that the more complex they are the less people will be able to understand the impact of it. Since the financial markets today are developed to the degree that orders can be executed in milliseconds model inaccuracies can cause unwanted effects on the market. Another aspect is whether it is necessary to create very complex models for something that can be estimated with a much simpler one or done without a model at all. This raises the question of whether it is enough with straightforward trading rules mentioned in previous section based on the basic idea of exploiting over- and undervalued stocks.
The next thing to consider is diminishing returns with increasing number of users of the same arbitrage strategy. David Shaw, founder of the successful hedge fund D.E. Shaw, expressed the trend of slimmer returns and that the success of his firm can be explained by its early entry.[5] Estimating the impact of this effect is however difficult.

[5] Gatev et al. (1999)

What is reasonable to assume though is that less popular pairs trading methods potentially should offer more opportunities than more popular ones where traders are competing for the same returns using the same pairs and strategies.

The stochastic control approach (SCA) is an example of quantitative pairs trading approach that is relatively new in the field. It is formulated as an optimization problem to maximize an investor's utility based on a model of the relationship between two stocks.[6] Mudchanatongsuk et al. applied the strategy on simulated price data based on the very same process used in the model which leaves the reader wondering whether it works as well in a real market environment.

What can be further improved to take the strategy one step closer to reality is to find a systematic and consistent method of choosing pairs to complement the strategy. Modeling the spread as a stationary stochastic process (Ornstein-Uhlenbeck) means that the spread is assumed to be mean-reverting. Spread stationarity is therefore a natural pair selection criterion and possibly with correlation as sub-criterion. The idea is that high spread stationarity combined with low correlation will be a good selection criterion for a pair. The high spread stationarity shows strong mean reverting behavior while the low correlation translates into many trading opportunities.

Regretfully, it is rare that spread stationarity has been shown to exist between any pair, particularly for an extended period of time. However, when looking at a shorter time span (1-3 years) in the absence of booms or recessions some level of stationarity does exist[7]. Transaction cost is a factor that is often ignored in academic papers. These can have a big impact on the performance of the strategy. Since the weight is updated for every time step the transaction costs can reach high levels which is why reduced trading frequency can be beneficial.

1.3 Purpose

The purpose of this thesis is to define a suitable pair selection method that supports the theoretical framework of SCA and to test this selection strategy against a more basic pairs trading strategy on historical market data.

1.4 Outline

The work will be organized in such way that hedge funds and pairs trading will be elaborated in chapter 2. Chapter 3 describes the methodology of this study. Practicalities such as research approach, data collection and trading simulation will

[6] Mudchanatongsuk et al. (2008)
[7] Wiktorsson, M., Associate Professor, mathematical statistics, Lund University (2010)

be described. Chapter 4 presents the practical framework of this study. Here, the trading simulation is presented and discussed, and practical issues concerning the trade are addressed. Chapter 5 describes the theoretical framework upon which this study is based. Mathematical theory is described in detail, as are the various selection methods employed in this study. Chapters 6, 7 and 8 show the results for pair selection, parameter selection and benchmarking. Chapter 9 discusses and compares the SCA.

1.5 Delimitations

A simulative method is used, enabling the study to become quite large very rapidly. Therefore, in order to avoid studying all possible combinations of screening strategies and parameter values, the study is organized as an optimization of a multivariable function.

In finance sector, trading strategies tend to be classified. It was therefore not possible to evaluate SCA against current strategies utilized by hedge funds. Instead, it is decided to compare to a basic pairs trading strategy, often found in the introductory section of text books on the subject.

2 Hedge funds & Pairs trading

2.1 Hedge funds

One of the big actors on the electronic stock markets is hedge funds who according to estimates manage assets in the range between \$1.8 and \$ 4 trillion.[8] Like mutual funds they manage portfolios that investors can invest in for a fee. This is where the differences between them begin. Unlike mutual funds hedge funds are less restricted and are allowed to short sell, use leverage and derivatives. Because of this flexibility hedge funds have a wide choice of investment strategies.

In the U.S. most hedge funds are subject to Regulation D 506 which only allows “accredited investors” to invest in them. Similar regulations can be seen in other countries as well. Examples of accredited investors include financial institutions, charitable organizations with assets exceeding \$5 million, high net worth individuals etc.[9]

As the name implies hedge funds are funds that protect their investments by hedging against loss risks at the same time as exploiting profit opportunities. Hedging against all risks will however hedge against returns as well. Therefore hedge funds only hedge against risks that carry potential losses while taking risks that are likely to pay off.[10]

The wide range of instruments and strategies used by hedge funds makes it hard to generalize them. One categorization used by Alternative Investment Management Association is shown in figure 1. The categories are ordered after increasing market exposure from left to right. Relative value strategies are sometimes called market neutral. This means that they are not correlated to how the market moves. Event driven strategies exploit opportunities caused by events like mergers and distressed companies. The last category, directional strategies are as the name implies strategies with a direction bias. They carry a higher market exposure than relative value strategies. It is not uncommon for hedge funds to use several investment styles to profit from different situations because some styles works better in some market situations.[11]

[8] Ineichen & Silberstein (2008)
[9] http://www.sec.gov/answers/accred.htm
[10] Ibid.
[11] Anonymous Stockholm-based hedge fund (2010)

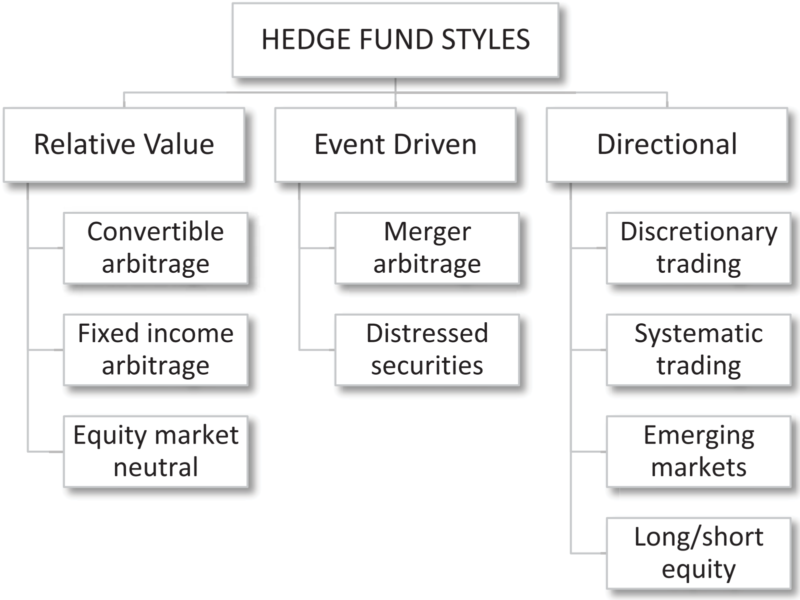

Figure 1 - Hedge fund styles with increasing market exposure from left to right[12]

Figure 1 shows the return pattern of a mutual fund against the Standard & Poor's 500 stock index (S&P 500). As can be seen the returns are correlated with the movement of the stock index S&P 500. The close correlation between mutual fund and S&P 500 returns can be explained by the restriction of short selling stocks. Figure 2 shows the market neutrality of a relative value hedge fund. No matter what direction the S&P 500 takes the returns are in general positive. This kind of strategy is also sometimes called absolute return strategies.

[12] Ineichen (2000), revised version in Ineichen (2008)

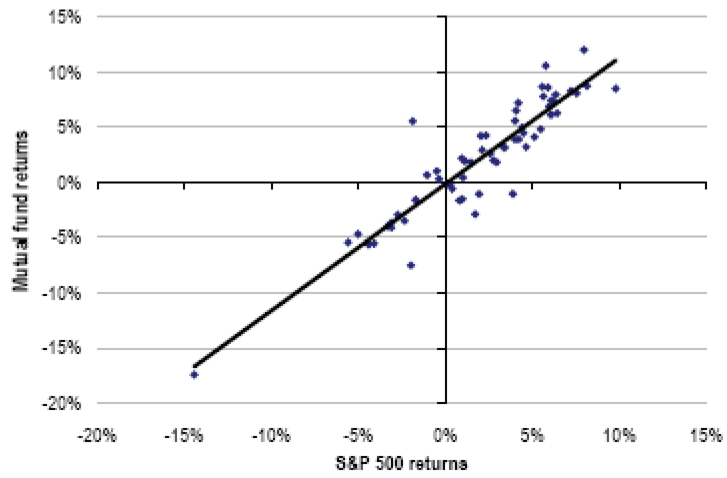

Figure 2 – Typical return pattern for mutual funds

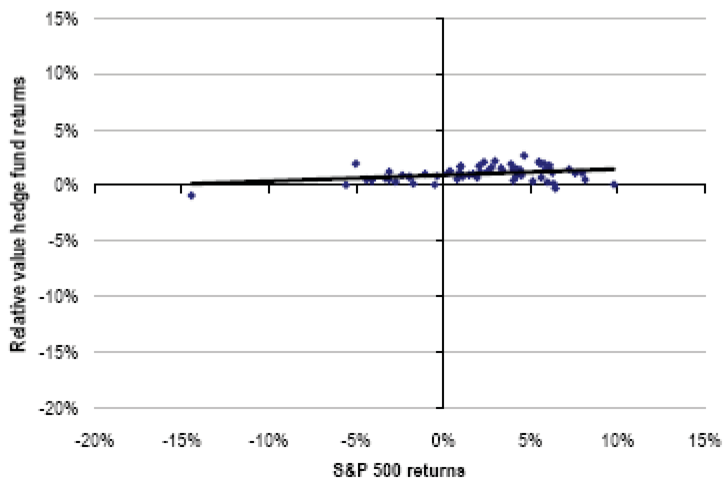

Figure 3 – Typical return pattern for relative value hedge funds

An important requirement for a hedge fund is consistent returns to ensure that clients receive a reliable return on their investment. A lot of work is done to find, test and implement strategies that fulfill demands on return volatility.[13] Depending on the character of the hedge fund different level of volatility is accepted. As shown in table 1 all hedge fund strategy styles offer higher annual return and lower volatility than MSCI World, an index of 1500 selected companies around the world.

[13] Anonymous Stockholm-based hedge fund (2010)

Table 1 Hedge fund performance characteristics (Jan 1990 - June 2008)[14]

Index	Annual return (%)	Volatility (%)	Sharpe ratio
MSCI World	7.2	13.9	0.20
JPM Gvt. Bonds	7.4	6.3	0.47
ML US 3M T-Bills	4.4	0.5	0.00
HFRI Fund Weighted Composite	13.3	6.7	1.34
HFRI Relative Value	11.4	3.6	1.94
HFRI Event-Driven	13.5	6.4	1.43
HFRI Equity Hedge	15.8	8.6	1.32
HFRI Macro	14.9	7.9	1.33

In terms of volatility relative value strategies offer the lowest volatility of the different strategy groups, shown in table 1. This goes well in line with the hedge funds serving volatility averse investors. One strategy under the relative value family is pairs trading which belongs to the sub-category equity market neutral.

2.2 Arbitrage and market efficiency

One important task of the stock market is to reflect proper values of companies for investors to make investment decisions. If stock prices correspond to the correct value of a company the market is efficient. As mentioned earlier pairs trading however depends on mispriced stocks within a certain time frame. Over- or undervalued stocks are shorted or bought in wait for price corrections. The cause of this mispricing is market inefficiencies. To go one step further one can say that it's caused by information asymmetry, which means that some investors are more informed than others.

The efficient-market hypothesis says that there are three levels of market efficiency: weak form, semi-strong form and strong form. The strong form states that all information, including information monopolized by a few investors is reflected in share prices. In this world arbitrage opportunities are close to non-existent. The semi-strong states that all public information like annual reports and news releases are reflected in share prices. The weak form covers only information about historical prices.[15] If the market is efficient in the strong form pairs trading would be very difficult to execute since prices are instantly corrected by information not available to everybody. At the same time it is not proven that market efficiency in its strong

[14] Ineichen (2008)
[15] Fama (1969)

form holds. There are different views on whether the semi-strong and weak form holds. [16]

Something to reflect upon is even though market is efficient there is the possibility of lag. An example of how this can be used is for a company listed on two different stock exchanges. Theoretically the price of the same company listed on different exchanges should be the same according to the law of one price. Otherwise it would be profitable to buy the stock in the exchange where it is priced lower and sell it in the other exchange. However the law of one price does not always hold because of asymmetry in information, market structure etc.

2.3 Categorization

Pairs trading is often seen implemented on equity strategies but it can be seen in other asset classes as well like convertibles (convertible arbitrage) and options (volatility arbitrage). Vidyamurthy classifies equity pairs trading into two categories: statistical arbitrage and risk arbitrage. Statistical arbitrage exploits price divergence as mentioned above. Risk arbitrage refers to a strategy that is used during mergers between two companies, one being acquirer and the other target. If the merger is successful the stock of the target will be converted into the acquirer stock and only one price will exist. The arbitrage opportunity exploits the often undervalued pre-merger rice of the target and overvalued price of the acquirer. The uncertainty with this strategy lies in the risk of the merger not being successful.[17]

Figure 4 shows the general work flow most pairs trading strategies. The two most important steps are choice of pair selection method and trading implementation. Trading implementation include strategy specific details such as choice of time frame, stop-losses, take-profits etc. Data collection and final pair selection can be seen as intermediary steps done on routine.

[16] Ibid.
[17] Vidyamurthy, 2004, pp. 8-9

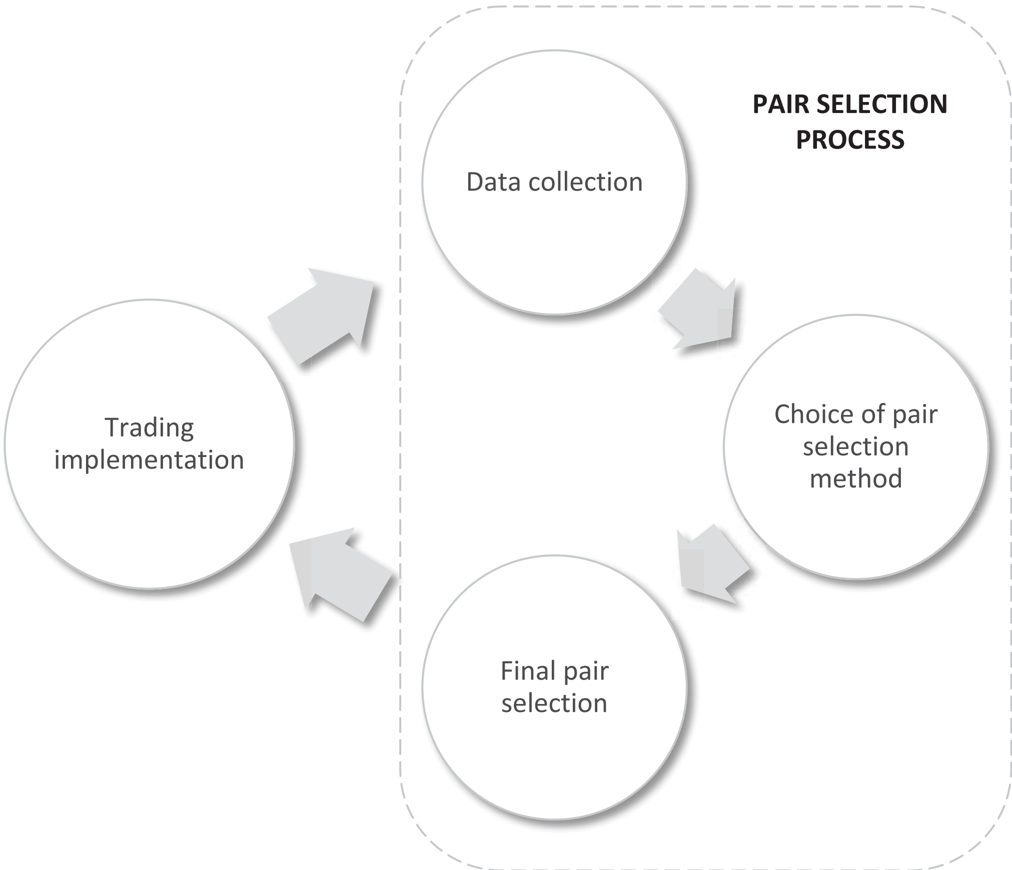

Figure 4 – General work flow for pairs trading

2.4 Pairs trading methods

Due to the proprietary nature of pairs trading few research papers have been written about it. In a paper from 2006, Do et al. mentions three main pairs trading methods: the distance method, the cointegration method and the stochastic spread method.[18] They all use slightly different methods in measuring the spread and strength of mean reversion. However, the main idea of entering a position when the mean reversion is strong enough and then unwind it when the spread is close to the mean is still a common treat in the different methods.

As an example, in the distance method a trading opportunity occurs when the distance crosses a trigger which is defined as a percentile of the empirical

[18] Do et al. (2006)

distribution. The distance is defined as the sum of squared differences between two normalized price series. There are three conditions that can close the trade: when the spread narrows to a certain limit, the last day of trading period is reached or when the distance increases further.[19] The distant method relies on the statistical relationship between the two stocks but have no forecasting ability.

More recent academic developments are the regime-shifting pairs trading rule[20] and stochastic control in pairs trading[21]. The focus of the former method is on identifying the shift of long-term equilibrium and defining trading rules accordingly. Mudchanatongsuk et al. on the other hand studies a different approach based on stochastic control (SCA). The spread is modeled with a Ornstein-Uhlenbeck process and formulate an optimization problem of the value function including the risk aversion of an investor. The optimized parameter is the weight of the stock pair in a portfolio that also contains a risk-free asset with a steady interest return.

[19] Nath (2003)
[20] Bock & Mestel (2008)
[21] Mudchanatongsuk et al. (2008)

3 Methodology

3.1 Research approach

In this paper a deductive research approach is used to study how well a quantitative stochastic control strategy works relative to a more basic pairs trading model. This means that empirics are derived from a theoretical starting point which in this study is the SCA.[22] Using this research approach there is the tendency of being limited in the search for data by the choice of theoretical framework. Therefore it is important to be careful in data selection to choose for the study relevant data.

3.2 Work flow

The many steps and procedures in this study are roughly illustrated in figure 5. Each step is then explained in further detail in the sub-sections.

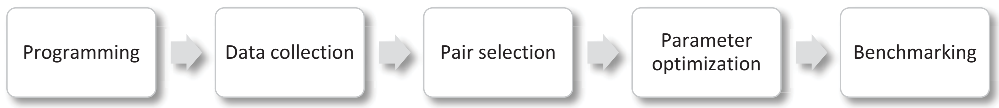

Figure 5 – Work flow chart.

The first step of this study is to create a program for trading simulation based on historical data (backtesting). Trading simulation is part of each of three last steps, pair selection, parameter optimization and benchmarking in figure 5.

3.2.1 Programming

The trading framework that this study is based on is programmed in Matlab. Further details regarding the program structure is described in chapter 4.1.

3.2.2 Data collection

One of the criteria used for selecting equities to trade is liquidity. High liquidity is necessary to keep bid/ask spreads down which is important for the strategy studied in this paper because of the frequent trading activity. Another issue regarding difficulties with finding lenders can arise as well if an illiquid stock is shorted.[23]

[22] Bryman & Bell (2005), p. 23
[23] Jacobs & Levy (1993)

A second criterion used considers the currency risk that companies bear. Although it is difficult to completely remove currency risk it can be lowered by only choosing stocks noted in the same currency. In this study Euro has been chosen. Companies in this study have been picked from the two largest Euro zone stock exchanges: Frankfurt Stock Exchange and Euronext.

Before starting to analyze the relationship on stock pairs it is necessary to narrow down the number of available stocks on the exchanges to reduce unnecessary work and save time for downloading data. In this study the first filtering of stock is based on liquidity and currency. The number of stocks from Frankfurt Stock Exchange and Euronext is limited to the 100 most liquid ones from each. The total number of companies involved in this study will therefore be 200. The secondary data used for pair selection and trading simulation is daily stock prices obtained from Datastream.

3.2.3 Pair selection

Different pair screening methods are tested in this study to see which one is more suitable for SCA. In this study two methods for testing for spread stationarity are used. The first one is where stationarity is tested for through the augmented Dicky-Fuller test (ADF test). In the second method, stationarity is examined through linear regression to determine whether this very simple test for stationarity is viable. Additionally, selection criteria based on correlation and industry belonging are examined. The mathematical details behind the methods are described in the chapter for theoretical framework. A negative control where final portfolio values are simulated on completely randomly chosen pairs is used. This negative control is used to compare each pair selection method to see how well pairs perform between the different selection methods and with a random choice of pairs.

Stationarity

All 19 900 pair combinations of the 200 companies are checked for stationarity with ADF test and linear regression which is performed in a Matlab programmed algorithm explained in section. The results are then put into a nx3 matrix, where n is number of pairs with stationary spread. The first two columns contain the indices for the two companies for each pair. The third column is the calculated correlation for which the matrix is sorted after. To see how well pairs perform depending on different stationarity periods a couple of periods are tested: 125 (half a year), 251 (one year) and 500 days (two years). To complement stationarity correlation is checked for pairs with stationary spread. The idea is to see whether stationarity by itself or in combination with correlation is a good pair selection criterion.

Correlation

The second relationship that is tested for is correlation. 64 random stocks have been chosen to form pairs. Correlation is then calculated for all 2016 pair combinations. These are sorted after correlation value $[-1, 1]$, and the final portfolio value is calculated accordingly. Then it is investigated whether there is a trend or connection between the correlation coefficient and profit.

Industry – the qualitative approach

Rather than focusing on stationarity, correlation or another quantitative aspect, this selection criterion is strictly qualitative. Instead of screening for a particular property, all stocks belonging to a particular industry are picked and possible pair combinations are made from these. For instance, if there are ten stocks among an industry, this would give 45 different pairs to test. Three industrial sectors are chosen: Oil & Gas, Health Care and General Retail. The selection was made on the basis of having as different industries as possible, and on the requirement of there being as many exchange noted companies within each industry as possible.

3.2.4 Parameter optimization

When a pair selection method is chosen, strategy parameters will be tuned to suit the SCA. There are five strategy parameters that are flexible to tune with but will stay fix throughout the trading period: gamma γ, trading period T, SCA parameter estimation period N and trading frequency ΔT. If stationarity tests or correlation is used for pair selection there is a fifth parameter, stationarity/correlation period T_C. These parameters should not be confused with the SCA parameters used for estimating σ, μ, κ, θ, η and ρ which will be explained in further detail in section 5.4. The SCA parameters are calculated automatically continuously during the trading period.

All strategy parameters except for γ are time frames. γ will be examined for the impact it has on final returns and the continuous portfolio value. This is a parameter which expresses the investor's risk aversion and through the formulas explained later the value of γ is assumed to only affect amplitude of return swings and not the direction of the trades. Therefore it is the first strategy parameter to be examined to confirm the hypothesis. The time frame parameters all have an impact on the final results but of different magnitude.

The first time frame is the one for stationarity test. It defines for how far back in time the pair spread needs to be stationary.

The second time frame is similar for the first one but used for estimation of SCA parameters σ, μ, κ, θ, η and ρ.

The third time frame is the trading period. This is the number of days the trading is active to achieve maximum utility as defined for SCA. As the strategy is formulated, it is this final value that is optimized in the model.

The final time frame is the transaction frequency period which is how often positions are updated. This parameter has an effect on transaction costs.

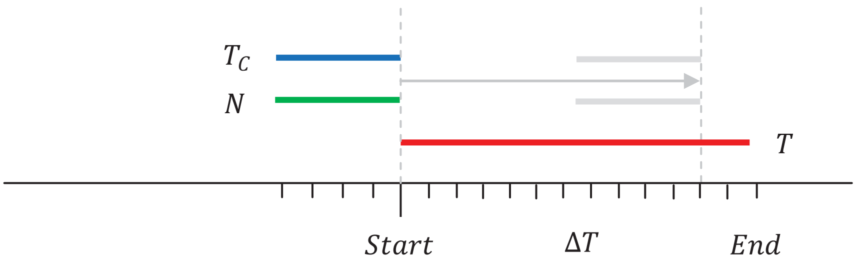

Figure 6 – Time frames illustrated

Illustration of the four time frames involved can be seen in figure 6. All are fixed in length for each simulation while T_C and N moves along the trading day.

3.2.5 Benchmarking

This is the final part of this study to see how much or if a sophisticated model is better than a pairs trading strategy with straightforward rules with no economic prediction capability. A well established and widely used pairs trading strategy is used as a benchmark to better illustrate the contrast between two different approaches to pairs trading. The basic pairs trading strategy used for benchmarking is one with threshold triggers for trade entries and exits. When the spread reaches two standard deviations above its mean a trade is entered.[24]

125 days is used to calculate the mean. No stop-losses have been used. The risk parameter used in the basic model is how much of the portfolio is risked on each trade. The position is unwinded when the spread reverts back to one standard deviation above its mean[25]. Only one pair position can be active at a time. The comparison between the two strategies is done for a trading period of one year,

[24] Gatev (1999)
[25] Ibid

each with a starting portfolio of €1000. The same set of pairs is used for both strategies as well to make the comparison valid. This means that the pairs used in the basic pairs trading strategy are screened for stationarity as well.

Finally the SCA coupled with the pair selection method chosen is benchmarked against a basic pairs trading strategy.

3.3 Data reliability

Data reliability is subject to the source from which data is collected. In our case it is the data obtained from Datastream which is a subscription based financial data service provided by Thomson Reuters. They are one of the biggest providers of financial data with Bloomberg being the other one. It is therefore considered to be a reliable source for accurate data. Restrictions with the Datastream student license used in this study are that the highest possible frequency of data is daily data and that there is a lag of data. Both restrictions do however not affect the study since the strategy is based on daily order updates and historical data is studied.

To further increase reliability of the study numbers are handled with care to avoid typos. Programs were tested with constructed matrices to ensure that indexing of price vectors from Datastream are done properly.

4 Practical framework

4.1 Programming

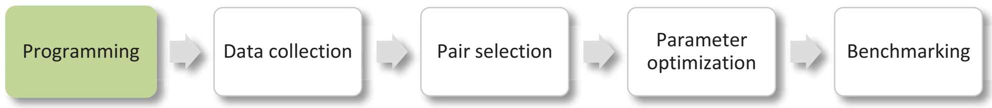

Figure 7 – Work flow chart.

The trading simulation programs are all written in Matlab. A schematic view of the program structure is illustrated in figure 8.

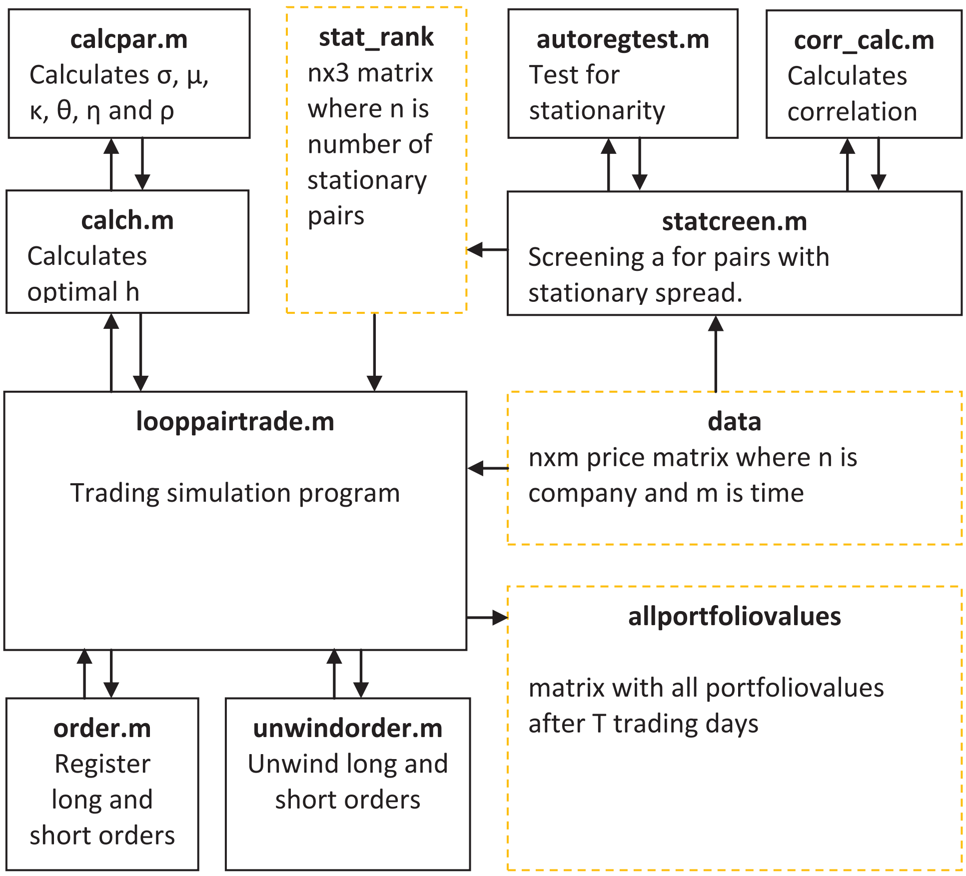

Figure 8 – A simplified structure of the trading simulation program. Dotted boxes represent matrixes and vectors, black boxes represent programs and functions.

During the trading simulation stationarity is tested continuously for the stationarity period with the purpose to detect fundamental changes in relationship between the stock pair. If no stationarity exists the weighting for each stock is set to zero, that is both positions are unwinded. The reason behind this is of precautionary nature. If there is no stationarity found the model is not valid and the impact of this is uncertain, therefore the positions are unwinded when this happens in this study.

A restriction was introduced on h not go above 1 or below -1 to avoid taking too large positions relative to the risk-free asset because it is not possible to lose more than the total investment liquidity. The restriction can also be seen as a requirement of a specific margin from the broker's side to cover the short position.

Portfolio value must be positive so trading is stopped if it reaches zero. The portfolio start value is set to €1000.

4.1.1 Programs

Here each program function from figure 8 is described.

autoregtest.m

This program tests stationarity on the spread between two time series that are obtained from the *data* matrix (see section 4.1.2) through *statscreen.m*.

corr_calc.m

This program calculates the correlation between two time series that are obtained from the *data* matrix (see section 4.1.2) through statscreen.m.

statscreen.m

Coordination of the stationarity testing is done in this program. Every possible combination of stock price vectors in the *data* matrix are sent to autoregtest.m for stationarity testing. All pairs that fulfill stationarity are then put into the *stat_rank* matrix where the pairs are sorted after correlation which is calculated through *corr_calc.m*.

calcpar.m

All strategy parameters σ, μ, κ, θ, η and ρ are estimated in this program based on formulas in appendix B.

calch.m

Through the estimated parameters in *calcpar.m* the optimal weighting for the long and short position is calculated in this program.

looppairtrade.m

This is the main program of the trading simulation that coordinates weighting calculation (*calch)*, order sizes (*order.*m) and portfolio value updating (*unwindorder.m*). Stock prices are obtained from the *data* matrix and information of which pairs should be simulated are obtained from *stat_rank*.

order.m

The daily order size is calculated daily in this program based on the optimal weight from *calch.m*

unwindorder.m

This program accounts daily profits or losses and accumulates it to the portfolio value.

4.1.2 Data matrices

data

This matrix contains all price data for the stocks in the study. The size of the matrix is nxm where n is number of companies and m is number of days included in the study.

stat_rank

This matrix with the size nx3 where n is the number of stationary pairs. The first two columns are the index of each stock in the pair in the *data* matrix. The third column is correlation. The matrix rows are sorted after correlation.

allportfoliovalues

This matrix contains all final portfolio values for the stationary pairs after T trading days.

4.2 Transaction costs

Transaction costs can be seen as a friction in trading. For the investor it is value that is lost in the investment process. It is a highly relevant issue for large investors who might incur market impact costs. Another group of investors who need to manage transaction costs are those who employ strategies with daily management of positions, especially within high frequency trading. The user of the SCA strategy is likely to fall into both categories. While not the focus of this paper it is important to understand and be aware of the impact of transaction costs. This section will explain different components of transaction costs and how it is dealt with in this study.

Two categories that transaction costs can be divided into are fixed and variable costs. Fixed costs include commissions and fees. As the name implies fixed costs are out of the investor's area of direct influence. Variable costs on the other hand are manageable to different degrees. These include taxes, spread, delay costs, price appreciation, market impact, timing risk and opportunity costs. Another dimension to categorize transaction costs is visible and non-transparent costs. Combining these two categorization dimensions creates following matrix in which the variable and non-transparent costs are the most manageable ones[26]:

Table 2 - Classification of transaction costs.

	Fixed	Variable
Visible	Commissions Fees	Taxes Spreads
Non-transparent	N/A	Delay costs Price appreciation Market impact Timing risk Opportunity costs

There is not much to comment on commissions and fees. They are usually fixed and nonnegotiable. Tax is an issue that is out of the focus of this trading strategy. When it comes to the spread the main determinant of its width is order imbalance. Order imbalance occurs when buy orders outnumber sell orders or vice versa. The inventory that the market maker has to hold overnight is a risk that is compensated by adjusting the bid-ask spread to attract orders. Whenever the inventory moves too far away from desired size the market maker will widen the spread to indirectly force orders to go towards the needed direction and thereby rebalance the

[26] Kissell et al. (2003)

inventory.[27] It is shown that the spread widens as trading volume rises.[28] This is consistent with the idea that investors can be divided into informed and uninformed. As example, during news releases when volume rises informed investors profits from the uninformed ones. The market maker need to balance the losses incurred from informed investors by widening the spread. Spreads for stocks tend to show a U-shaped pattern which means that the spread is wider in the beginning and toward the end of the trading day than mid-day.[29] The data for this study is based on close price which means that the spread costs are overestimated if orders are put in towards mid-day in the real implementation of the strategy.

Delay costs arise from the delay between investment decision and the moment the investment reaches the market. The investment workflow can be improved for manual trade setups while for automatic algorithm trading setups delay costs are already small.

Market impact refers to the impact an order has on the market price. It is impossible to tell exactly how much an order pushes the market price equilibrium since price development can't be measured both with and without the order. What an investor can do instead is to observe historical price movement patterns caused by its orders to obtain a rough estimation of market impact. With that knowledge a model can be developed to slice up the order into smaller pieces that are released to the market with certain time intervals.[30]

Because big orders are broken into smaller ones to avoid too big market impact the price will change between each order. This effect belongs to the price appreciation category in the table above. Opportunity costs arise from illiquidity and adverse price movements during trading.[31]

Total transaction costs can take up 10-20% of order value. Experience is important when it comes to estimation of transaction costs. [32] A big part of this figure is already accounted for in the close prices used in this study, i.e. delay costs, market impact etc. The optimization of transaction costs is often done by a dedicated group within a hedge fund. Costs that are not accounted for are commissions and fees which usually are not a big part of total transaction costs and can vary a lot between brokers and market makers.

---

[27] Chan et al. (1995)
[28] Lee et al. (1993)
[29] Chan et al. (1995)
[30] Anonymous Stockholm-based hedge fund (2010)
[31] Kissell et al. (2003)
[32] Anonymous Stockholm-based hedge fund (2010)

4.3 Short selling

Short selling is the practice where a stock is sold that is not owned by the seller. What happens is that the seller borrows the shares from a lender for a fee and sells them in the market for the actual market price in hope that the price will fall. The position is closed when the short seller buys back the shares from the market and returns them to the lender. If the share price has fallen during the short selling period the short seller profits from the price difference minus lending fee and other transaction costs. Short selling can cause unwanted amplification of bearish trends as during the recent financial crisis. Some countries banned short selling temporary during the crisis to let share prices re-establish. U.S. Securities and Exchange Commission (SEC) for example introduced a temporary restriction on short selling of 799 companies September 2008.[33] Financial market regulators in many other countries like U.K.[34], Germany[35], France[36] and Australia[37] introduced a temporary ban at the same time as well.

The seller is usually given a certain time frame to deliver the shares that are sold to the buyer if they are not already secured from a broker before the transaction. In the U.S. Regulation SHO is a regulation founded by SEC 2005 with the purpose to prevent unethical short selling. It uses the common settlement cycle of three days, also known as T+3. This means that transactions should be completed within three business days.

The seller is obligated to deliver the stocks and the buyer is obligated to pay within the time frame. The regulation prohibits the broker from contracting to settle transactions later than T+3. It is however not a violation if “fail to deliver” occurs.[38] Failure to deliver can sometimes occur of technical reasons with no unethical intentions which is why the regulation is formulated in such way. Both the German and French financial market regulators BaFin and AMF have a regulation similar to that of SEC with T+3.

One version of the short selling described above is naked short selling. The difference is that shares are not borrowed or checked for availability for the shorted stock before shorting them. What can happen in an unregulated market is that naked short sell are kept in the book while stocks are waited to be borrowed and

[33] SEC (2008-09-19)
[34] FSA (2008-09-18)
[35] BaFin (2008-09-19)
[36] AMF (2008-09-19)
[37] ASIC (2008-09-21)
[38] SEC Regulation SHO (2005)

thus artificially pushing the share price down. To prevent this regulating agencies try to impose regulations for closing failure to deliver positions, like the one found in Regulation SHO.[39]

Short selling regulations change constantly, especially during the volatile period during the recent financial crisis. This puts more restrictions on the practitioner and flexibility is necessary for liquidity. Often, markets force investors to be able to cover their short selling risks by not betting more than they afford to lose, meaning the short selling cannot be greater than the total portfolio value. This validates the restriction set in 4.1.3 further.

[39] Ibid.

5 Theoretical framework

Here the mathematical theory required for the simulation is presented. First, the fundamentals of stochastic stock modeling is presented and explained how it is applied to the SCA. Then the mathematical basics of subsequent s selection criteria are described.

5.1 Stochastic Control Approach

There are several assumptions made concerning the nature of changes in the stock prices. It is assumed that the stock price follows a stochastic process where the stock follows a lognormal distribution. Further, it is assumed that the changes in stock price over time can be described with[40]

$$dB(t) = \mu B(t)dt + \sigma B(t)dZ(t) \qquad (1)$$

where B(t) is the stock price, μ is the drift, σ is the volatility and Z(t) is the standard Brownian motion. Brownian motion is a simple assumption stemming from particle physics suggesting that particles in suspension move in a completely random/stochastic way. More generally, the Brownian motion can be substituted with the Wiener process, defined as the pattern a stock price follows if it is subject to the Markov property (the future position is only dependent on the present and has nothing to do with the past). The change in the price of stock B(t) is subject to two forces; constant drift and random motion. The size of the change is in proportion to the price at each moment.[41]

Next, rather than analogously modeling the stock price for the other stock in the pair, A(t), the spread is modeled between the two as

$$X(t) = \ln\big(A(t)\big) - \ln\big(B(t)\big) \qquad (2)$$

Taking the natural logarithm of a stock series renders them normally distributed, a requirement that needs to be fulfilled in order to obtain a closed formed solution. And since both stocks are log-normally distributed, X will follow a standard normal distribution. X(t) is assumed to be mean reverting (i.e. there is a fix long term equilibrium value to which the spread will always try to revert) and is therefore modeled with the classic Ornstein-Uhlenbeck process as[42]

[40] Hull (2005), p. 265
[41] Ibid. p. 265
[42] Øksendal (2000), pp. 61-62

$$dX(t) = k\big(\theta - X(t)\big)dt + \eta dW(t) \qquad (3)$$

where $k \cdot \big(\theta - X(t)\big)$ is the drift term, with a constant rate of reversion $k > 0$ and the long-term equilibrium θ to which the process reverts to. η and dW(t) are standard deviation and Brownian motion for the spread.

5.1.1 Ito's lemma[43]

Ito's lemma is the equivalent, although seemingly more complex version of the ordinary derivate chain rule for deterministic functions but applied to stochastic processes. The theorem stipulates that if taking a variable x subjected to the Markov property, i.e.

$$dx = a(x,t)dt + b(x,t)dz \qquad (4)$$

where a and b are functions of x and t and dz is the Wiener process and/or Brownian motion, a new function G(x,t) will follow the process

$$dG = \left(\frac{\partial G}{\partial x}a + \frac{\partial G}{dt} + \frac{1}{2}\frac{\partial^2 G}{\partial x^2}b^2\right)dt + \frac{\partial G}{\partial x}bdz \qquad (5)$$

where dz is the exact same process as in x.

5.1.2 Deriving A

Taking the two equations for B and X and applying Ito's lemma to them, allows us to derive A (dA is the equivalent of dG in Ito's lemma) as

$$dA(t) = \left(\mu + k\big(\theta - X(t)\big) + \frac{1}{2}\eta^2 + \rho\sigma\eta\right)A(t)dt + \sigma A(t)dZ(t)$$

$$+\eta A(t)dW(t) \qquad (6)$$

The emerging term $\rho \cdot \sigma \cdot \eta$ (ρ being the direct correlation coefficient between the Wiener processes) is describing the covariance between the two stochastic processes Z(t) and W(t).

5.1.3 Portfolio value[44]

The trading strategy utilized and the limitation caused by the stock pair behavior modeling leads to the construction of a market neutral self-financing portfolio with a fixed amount of wealth in the risk-free asset and equal weighting for the two stocks.

[43] Kioyshi (1951), pp. 1-51
[44] Mudchanatongsuk et al. (2008)

The risk-free asset can function as collateral if required in the market the investor is trading in. Because both stocks carry equal weighting it means that if the portfolio is long $40 dollars on stock A, it must be short $40 on stock B. Put in mathematical terms this means the weighting functions have the relationship

$$h_A(t) = -h_B(t) \qquad (7)$$

In a stock portfolio, the sum of all individual weighting is always 1;

$$x \in \boldsymbol{R}^n$$
$$\sum_1^n x_i = 1 \qquad (8)$$

Since $h_A(t) = -h_B(t)$, the weight on the risk-free asset is always 1. Thus the change in the value of the portfolio is given by

$$dV(t) = V(t)\left(h_A(t)\frac{dA(t)}{A(t)} + h_B(t)\frac{dB(t)}{B(t)} + \frac{dM(t)}{M(t)}\right) \qquad (9)$$

Taking all of the equations together, this equation can be rewritten to include the stochastic processes described above

$$dV(t) = V(t)\left(h_A(t)\left(k\big(\theta - X(t)\big) + \frac{1}{2}\eta^2 + \rho\sigma\eta\right) + r\right)dt$$
$$+ h_A(t)\eta dW(t) \qquad (10)$$

The derived expression of portfolio value can then be inserted into the utility function

$$U(x) = \frac{1}{\gamma}(x)^\gamma \qquad (11)$$

where γ is the investor's risk aversion. The problem can now be defined as a maximization problem of the utility function. To maximize the function it is necessary to define the day to which the portfolio should be maximized. As a result of defining the end day T the trading period is also determined. The optimization problem is thus formulated as

$$\sup_{h(t)} E\left[\frac{1}{\gamma}\big(V(T)\big)^\gamma\right] \qquad (12)$$

Solving this will give the optimal weight $h^*(t)$ which determines how much and whether to long or short on each stock.

5.1.4 Derivation of the weighting function h[45]

To solve this problem the real function of three variables $G(t, v, x)$ is set to be the value function, and using the Hamilton-Jacobi-Bellman equation and the supremum criteria $h_A(t)$ can be derived as

$$h_A(t) = -\frac{\eta^2 \frac{\partial^2 G}{\partial v \, \partial x} + b \frac{\partial G}{\partial v}}{\eta^2 v \frac{\partial^2 G}{\partial v^2}} \qquad (13)$$

$$\text{where } b = -k(x - \theta) + \frac{1}{2}\eta^2 + \rho\sigma\eta$$

In an attempt to stipulate G, different solution approaches can be used, reducing G to a product of one variable functions. Eventually, under our conditions, it is possible to deterministically derive $h_A(t)$ as

$$h_A(t) = \frac{1}{1-\gamma}\left[\beta(t) + 2x(t)\alpha(t) - \frac{k(x-\theta)}{\eta^2} + \frac{\rho\sigma}{\eta} + \frac{1}{2}\right] \qquad (14)$$

Formulas for calculating $\alpha(t)$ and $\beta(t)$ are disclosed in Appendix C.

5.2 Integration and cointegration

Generally, a series is said to be integrated if there exists some sort of relationship (linear, polynomial, sinusoidal, exponential, logarithmic etc.) recursively among its elements that makes the series stationary[46]. The most simplistic example would be a vector

$$x = (1, 2, 3, 4, 5)$$

Differencing x once gives us the vector (note the dimension reduction)

$$\Delta x = (1, 1, 1, 1)$$

Thus, vector x is said to be linearly integrated of order 1, denoted $I(1)$. The order implies that it is necessary to difference (or derive) the vector n order of times, n-being the order, before the series becomes stationary. For instance, the vector

[45] Mudchanatongsuk et al. (2008)
[46] Granger (2004)

$$x = (1, 4, 9, 16, 25, 36, 49)$$

is linearly integrated of order 2, $I(2)$.

In practice, it is more interesting to study integration among different vectors and time series[47]. This phenomenon is called cointegration. For this study, only pairs will be examined, which is why the theory is limited to cointegration of two vectors. In the case of two vectors, differencing or finding the derivate becomes less important, rather, some combination of independently non-stationary vectors is looked after that upon some operation become stationary. Consider the simple system of stochastic equations given by:

$$\begin{cases} y_{1t} = \alpha y_{2t} + \varepsilon_{1t} \\ y_{2t} = y_{2,t-1} + \varepsilon_{2t} \end{cases} \qquad (15)$$

where $\varepsilon_i \in N(0,1)$. y_{2t} is a random walk process, i.e. I(1), and y_{1t} is scaled y_{2t} with the addition of another stochastic variable. Since the sum or normal distributions is a normal distribution, y_{1t} is also I(1). Both series being independently integrated of order 1 , I(1), is a requirement for eventual subsequent cointegration. Rearranging the first equation gives us

$$y_{1t} - \alpha y_{2t} = \varepsilon_{1t} \qquad (16)$$

ε_{1t} is stationary with the mean 0, meaning that the two processes y_{1t} and y_{2t} are linearly cointegrated. To denote the cointegration relationship between two variables, a cointegration vector is used. The cointegration vector can be viewed as the equivalent of the I-operator ($I(0)$, $I(1)$, etc) for several variables. In the case above, one cointegration vector would be

$$(1, -\alpha)$$

Naturally, the vector

$$(-3, 3\alpha)$$

is a cointegrating vector as well, but they will only be called different cointegration vectors if they are linearly independent. So any vector $X = [x_1, x_2]$, which makes the product [48]

$$X \cdot \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}$$

[47] Hamilton (1994), p. 572
[48] Hamilton (1994), pp. 572-574

stationary, is a cointegrating vector. If no such vectors can be found, y_1 and y_2 are not cointegrated. More generally, cointegration can also refer to integration of more than two variables and of different orders. But this multicointegration is not used in our study.

Today, two main tests for regular cointegration exist. The Engle-Granger method[49] uses a two step statistical analysis determining if the residuals follow a random walk or not. This method is somewhat limited, but is very useful for solving problems of minor complexity. The Johansen test[50] is a more generalized version of the Engle-Granger solution; for instance, it is possible to detect different orders of cointegration with this method. There exist slight variations on each method, such as the Phillips-Ouliaris cointegration test.

However, for this study, the requirement for cointegration is too strict. The is no demand of the stocks being integrated of order 1, or rather their natural logarithm being integrated of order 1, although it could be. There is also no requirement of finding some relationship (i.e. some cointegrating vector) that renders a combination of the two series stationary. Rather, the requirement is that the logarithmic spread is stationary (and mean-reverting), which can be interpreted as a very special and unusual type of cointegration. The behavior of individual stocks does not interest us per se. That is the fundamental of a market neutral strategy.

5.2.1 Calculating stationarity

In the closed form solution to the optimization problem, it is assumed that the logarithmic spread is stationary, or more specifically mean-reverting

$$X(t) = \ln(A) - \ln(B) \qquad (17)$$

$$dX(t) = k\big(\theta - X(t)\big)dt + \eta dW(t) \qquad (18)$$

In order to verify this, two different tests for stationarity. The first, and the more basic one will be a linear regression, using ordinary least square minimization. The spread will be modeled after:

$$X(t) = \alpha + \beta t + \varepsilon_t \qquad (19)$$

If β can be verified to be 0, the spread is assumed to be stationary. The downfall with this simple method is that based on the structure of the test (as will be

[49] Engle & Granger (1987)
[50] Johansen (1991)

described below), it can only tested, with a certain level of certainty that the spread is not stationary. However, eliminating pairs from a sample that are not stationary does not imply that the rest are.

The second test is the more intricate augmented Dickey-Fuller test (ADF). Originally, this is a stationarity test, a part of the more intricate Engle-Granger test for cointegration. However, in this case, it is enough just to test for stationary, not delving into whether the stocks are cointegrated or not. Here, it is investigated whether the spread follows a random walk or not. If not, then the process is stationary.

5.2.1.1 Linear regression[51]

Statistically testing whether β is zero in

$$X(t) = \alpha + \beta t + \varepsilon_t \qquad (20)$$

when fitting the data to straight line will imply that there is no relation between X and t, indicating that the spread is constant over time.

The estimated β-parameter in the regression line is given by $\beta = \frac{S_{tx}}{S_{tt}}$ where the three different sums are

$$S_{tt} = \sum_{1}^{n} (t_i - \bar{t}) \qquad (21)$$

$$S_{tx} = \sum_{1}^{n} (t_i - \bar{t}) \cdot (y_i - \bar{y}) \qquad (22)$$

$$S_{xx} = \sum_{1}^{n} (x_i - \bar{x}) \qquad (23)$$

The variance of this estimation is equal to

$$s^2 = \frac{Q_0}{(n-2)} \qquad (24)$$

where Q_0 is the difference/error between the actual values of X and the values of X calculated through the linear regression for every value t squared, thus

[51] Groβ (2003), ch. 2

$$Q_0 = \sum_{1}^{n} (x_i - \alpha - \beta t_i)^2 \qquad (25)$$

Using the three different sums this relation can be rewritten as

$$\mathrm{Q}_0 = \mathrm{S}_{\mathrm{xx}} - \frac{\mathrm{S}_{\mathrm{tx}}^2}{\mathrm{S}_{\mathrm{tt}}} \qquad (26)$$

making the calculations much easier. The regression is aimed at estimating two parameters α and β, meaning that the degrees of freedom need to be reduced by two $(n-2)$.
The verification whether cointegration exists is finally verified through a standard hypothesis test.

$$\begin{aligned} H_0 &: \beta = 0 \\ H_1 &: \beta \neq 0 \end{aligned} \qquad (27)$$

The estimation of the parameters in the linear regression are distributed after t-distribution, and when testing for 5% significance, both positive and negative values have to be considered, making the $t_{0.025}$ quintal our reference value.

The statistical unit $T = \frac{\beta - \beta_0}{d(\beta)}$ (where $\beta_0$is always 0) is compared to the $t_{0.025}$ quantil. $d(\beta)$, the standard error is equal to $\frac{s}{\sqrt{S_{tt}}}$. If $|T| < t_{0.025}$ it is concluded that H_0 is true. Otherwise, H_0 is discarded in favor of H_1 and next pair is examined.

5.2.1.2 Dickey-Fuller[52]

In the Dickey-Fuller test, the spread is modeled with the autoregressive equation:

$$X(t) = \phi \cdot X(t-1) + \varepsilon_t \qquad (28)$$

The hypothesis test of interest then becomes

$$\begin{gathered} H_0: \phi = 1\ (unit\ root) \rightarrow X(t) \sim I(1) \\ H_1: \beta < 1 \rightarrow X(t) \sim I(0) \end{gathered} \qquad (29)$$

$\phi > 1$ is not assumed to be possible, since this will cause the spread to be explosive, which is clearly not the case. Therefore, the t-test performed here is one-sided. ϕ is estimated in the same way as β in the test above, and the test statistic is analogously

[52] Dickey, D., Fuller, W. (1979)

$$t_{\phi=1} = \frac{\hat{\phi} - 1}{d(\hat{\phi})} \qquad (30)$$

$d(\hat{\phi})$ being the regular standard error estimate. The Dickey-Fuller distribution density is slightly skewed to the left from the standard normal, so the quantiles simulated as in[53], will be used. If $|T| > t_{0.05}$, H_0 is discarded in favor of H_1, and the is assumed to be stationary.

5.3 Correlation

Just like stationarity, there are several ways of calculating the correlation for two time series as well. Here, the most common one is used; Pearson product-moment correlation coefficient. Given two random variables x and y, correlation between these two is defined as the covariance divided by the product of the individual standard deviations. Put in mathematical terms this equals to

$$r_{xy} = \frac{cov(x,y)}{\sigma_x \sigma_y} = \frac{\sum_{i=1}^{N}(x_i - \bar{x}) \cdot (y_i - \bar{y})}{(n-1) \cdot s_x \cdot s_y} \quad , \qquad 1 \leq r_{xy} \leq 1 \qquad (31)$$

where r_{xy} is the correlation coefficient, $\bar{x}$ and $\bar{y}$ are the mean values of each vector respectively, and s is the intermutual standard deviation.

[53] MacKinnon (1996)

5.4 Parameter estimation[54]

To calculate the value function $h(t)$ six parameters have to be evaluated beforehand. Each parameter is derived through a mathematical function of two series of data, the spread, X(t), and the second stock, B(t). At every time point (day), the parameters become re-evaluated based on the new stock data that becomes available. The parameters are

- σ = the volatility of B
- μ = the drift of B
- κ = rate of reversion towards the mean of the spread
- θ = long-term equilibrium of the spread
- η = the standard deviation of the spread
- ρ = correlation coefficient between the Brownian motions for the B-stock and the spread X.

For the formula of for each parameter, please refer to Appendix A.

[54] Mudchanatongsuk et al. (2008)

6 Results – Data collection and pair selection

6.1 Data collection

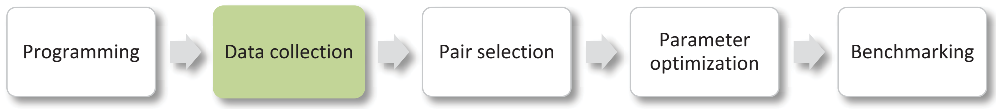

Figure 9 - Work flow chart

Two main criteria are used for selection of trading equity. The first one is liquidity. High liquidity is necessary to keep bid/ask spreads down which is important for the strategy studied in this paper because of the frequent trading activity. Another issue regarding difficulties with finding lenders can arise as well if an illiquid stock is shorted. The second criterion used considers the currency risk that companies bear. Although it is difficult to completely remove currency risk it can be lowered by only choosing stocks noted in the same currency. In this study Euro has been chosen. Companies in this study have been picked from the two largest Euro zone stock exchanges: Frankfurt Stock Exchange and Euronext.

Before starting to analyze the relationship of stock pairs, the number of available stocks on the exchanges were narrowed down to reduce unnecessary work and save time for downloading data. In this study the first filtering of stock is based on liquidity and currency. To obtain a sufficient number of pairs, the number of stocks from Frankfurt Stock Exchange and Euronext is limited to the 100 most liquid ones from each. The total number of companies involved in this study will therefore be 200. The secondary data used for pair selection and trading simulation is daily stock prices obtained from Datastream. Daily data from 1996 (the year Euro was introduced) to 2010 were collected to guard against change in time frames.

Stock market indices are not constant because companies enter and leave. Reasons for entering can be initial public offerings and transfer from other indices. Reasons for leaving can be mergers, acquisitions and share buy backs. This causes discontinuity in data and some companies will naturally fall out of the study in different time frames because of mentioned reasons. This will however not affect this study since a time period is chosen where all 200 companies are present continuously. The starting time for all analyses in this study is based on 1 January 2008. The choice of the starting date is to make sure that the market neutral strategy holds since 2008 is a strong bearish year.

6.2 Pair selection

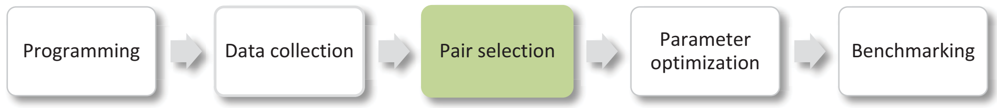

Figure 10 - Work flow chart

Next, the effects of different selection methods are studied individually. We describe the distributions of portfolio values for pairs that have been selected through 2 general methods; quantitative, and qualitative. The quantitative methods include stationarity through linear regression, stationarity using the ADF test, and correlation (thus automatically selecting the pairs of interest among the 19 900 possible pairs). The qualitative method resorts to sorting stocks after industries they belong to. Each simulation is compared to other simulations within the same category, i.e. different methods for stationarity are compared to one another, and to the negative control. The negative control is a simulation run completely random pairs, chosen stochastically from the stock data base consisting of 19 900 pairs. Conclusively, the optimized portfolio is compared to the alternative pairs trading strategy.

6.2.1 Key figures

In the evaluation, six key figures will be used. We look at the mean, median, standard deviation, maximum, minimum and data points for the terminal values of different stock pairs. For all practical purposes, it is the mean value that is the most critical one, since the goal is to develop better-than-average performance for a group of different stock pairs based on various selection methods. The starting value of each investment is €1000, and the higher number is achieved, the better.

The median value is not as critical as the mean, but gives, together with the standard deviation, an indication of how disperse the different terminal values are. The minimum and maximum values show the end points of the spread, and are an additional complement to the description of the series. Overall, low standard deviation is preferable, as is small a difference between the maximum and minimum values while maintaining a high mean.

The data points is a measurement of how many pairs among the 19 900 possible fitted in the particular selection criterion. This optimum would be to have fairly few data points (i.e. pairs) while having a substantially elevated mean value. This would indicate that a successful screening strategy able to select just dozens of pairs among thousands of possible ones has been developed.

6.2.2 Post scriptum program modifications

Upon initial simulations, one shortcoming in the algorithms was noticed. Upon parameter calculation, differences in S, ($S = \ln B$) has to be evaluated for different time points and use in the nominator (see Appendix B). If S is constant in this interval, infinity upon division by zero is obtained since both Δs_t and $\widehat{m}$ become zero with constant B. This renders the weighting function h(t) infinite as well. To go around this problem, the h(t) value is set to zero in case there is no difference in the logarithmic price of stock B.

6.2.3 Stationarity

Here, the two stationarity test methods; using the β-value and the ADF test are presented.

Stationarity through linear regression

The mean and the median values of the pairs that were stationary with the linear regression model (table 3) had a difference of €15 compared to the control (section 6.6). This difference, however, only caused the mean and the median in the linear regression model to retain the initial portfolio value. Although the mean of portfolio values is equal to the starting value, the stocks with stationary spread offer over 3 times less volatility, as measured by the standard deviation. The minimum and maximum values show less variation as well. These two values are not evenly distributed around the mean, although the median remains at the initial value. It is inconclusive to view just these numbers, but looking at the spread graph it is observed that there are approximately an equal number of profitable pairs as unprofitable. The difference in the unprofitable end appears to be caused by a small number of pairs with intense (20 to 25%) drop in value over the time period. Because of decreased volatility, the pairs with stationary spread are more suitable for usage in the hedge fund strategies. In some way, stationarity is clearly a desired property.

Table 3 - Several statistical parameters for the final portfolio values for stock pairs whose spread was stationary through the linear regression test.

Mean	1000
Median	1000
Std	25
Max	1192
Min	758
Data points	2005

Stationarity through Augmented Dickey-Fuller test

The mean of the sum of portfolios of pairs separated after ADF stationarity (table 4) is €3 above the starting value and €18 over the control. A significant difference is observed in the maximum and minimum values spread compared to the control. The distribution around the mean for these values is inverted if compared to the linear regression model with the median remaining at starting value 78 (figure 11). We are also able to attain higher mean revenue with half the standard deviation compared to the control. This property, combined with a not too low minimum value, makes us select this selection method for deeper study. Note that there were 40% lower amount (1236 versus 2016) of pairs fulfilling the stricter ADF requirement of stationarity compared to the simpler linear regression model.

Table 4 - Several statistical parameters for the final portfolio values for stock pairs whose spread is stationary through the ADF test.

Mean	1003
Median	1000
Std	46
Max	1677
Min	666
Data points	1236

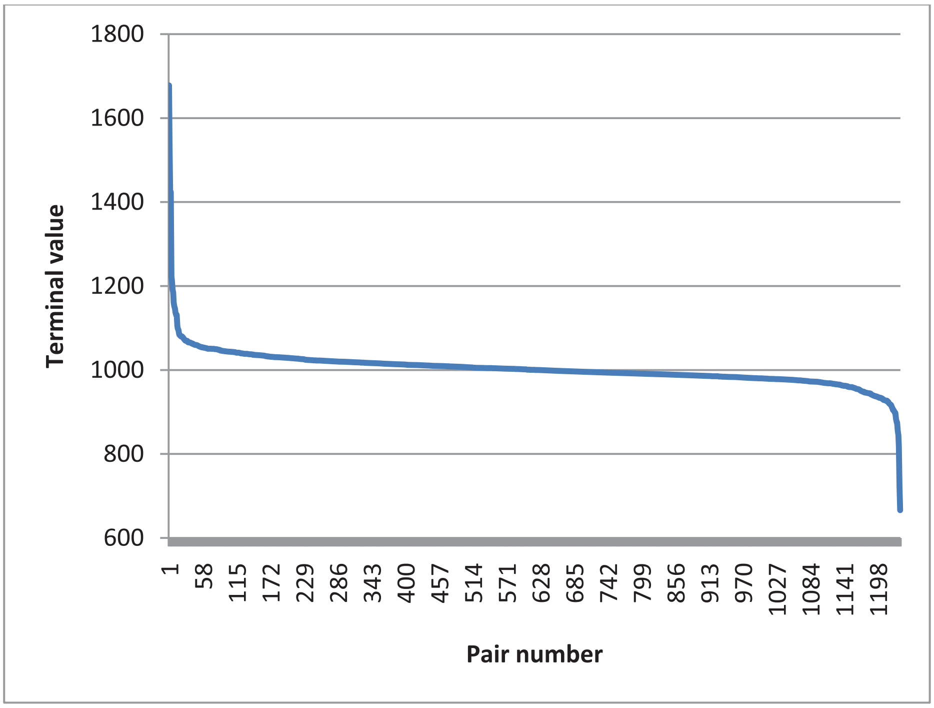

Figure 11 – An illustration of the result after a simulation. In this particular simulation the stock pairs were screened for stationarity using the ADF-test and the simulations were run for 125 days. The portfolios were sorted after their terminal value, with the one with the highest terminal value getting assigned the number of one, and then continuing in descending order. The number of the portfolio is on the x-axis and its terminal value is on the y-axis. Each pair was positive in the ADF stationarity screen.

6.2.4 Correlation

Pairs with high correlation have marginally higher revenue than the rest, although peeks and dips of various sizes are present throughout the entire correlation axis (figure 12). Although pairs with high correlation show higher revenue on occasion, it is noted that it is within the high correlated pairs that the lowest overall values are found as well. The divergence of the mean towards lower and higher value as correlation increases makes us omit the selection criteria based on correlation.

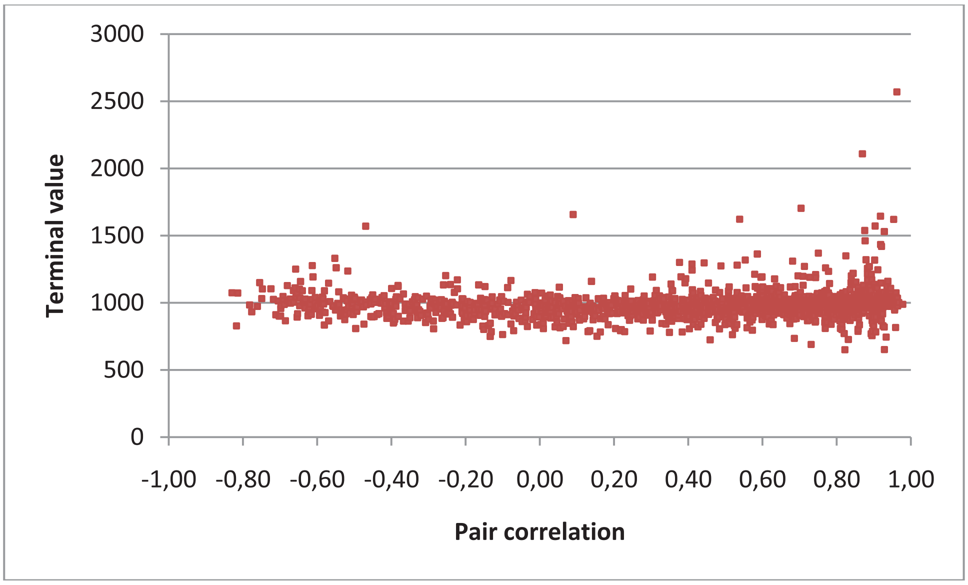

Figure 12 - Terminal values scattered after the correlation of their respective stock pairs.

6.2.5 Industry

Oil & gas

The mean value of this industry is almost identical to the control (table 5). Although the max is larger, the mean is smaller, implicating a larger standard deviation.

Table 5 - Several statistical parameters for the final portfolio values for stock pairs from the oil & gas industry.

Mean	985
Median	998
Std	161
Max	2691
Min	322
Data points	325

Healthcare

The healthcare industry is below Oil & gas, showing a larger standard deviation but all the other values being lower (table 6).

Table 6 - Several statistical parameters for the final portfolio values for stock pairs from the healthcare industry.

Mean	949
Median	982
Std	165
Max	1432
Min	483
Data points	253

General retail

General retail offers a comparatively impressing mean, although the median shows there to be more unprofitable pairs (table 7). Even so, these two values are above the control. However, the standard deviation and the difference between maximum and minimum values are so great, and only good for one out of three industries that the industry selection method is not chosen as a viable strategy. Without setting the restriction of trade termination if the portfolio reaches the value of zero, two pairs went as low as -1500.

Table 7 - Several statistical parameters for the final portfolio values for stock pairs from the general retail industry.

Mean	1008
Median	992
Std	178
Max	1790
Min	0
Data points	276

6.2.6 Randomly chosen pairs – negative control

In the negative control, approximately 2000 pairs were randomly selected from the possible 19 900. These were run in the simulation program with no sub-criteria or restrictions. The mean and median values are €15 below the initial value (table 8 and figure 13). The maximum value is over double of the initial, and there are several other pairs that achieve this high result. Clearly, there are some stock properties that account for this effect.

Table 8 - Several statistical parameters for the final portfolio values for 2016 randomly selected pairs.

Mean	985
Median	984
Std	85
Max	2109
Min	543
Data points	2016

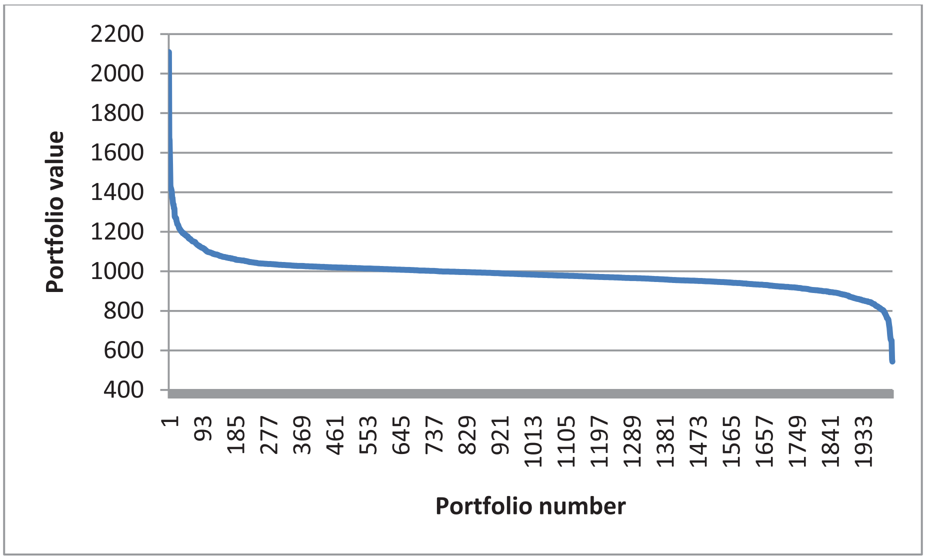

Figure 13 – The spread of terminal portfolio values for the 2016 randomly selected stock pairs.

6.2.7 Results subject to further study

Based on these results, and the overall requirement of hedge funds of having a stable portfolio spread, it was determined that the ADF stationarity method was the only good enough initial selection criteria worth of further investigation. However, before continuing on with the parameters optimization study, a test was made to verify that it was indeed the logarithmic stationarity that was essential, and not just any other.

Sporadically, we choose to test this with the most common form of stationarity, the linear, table 9. The difference in mean value for the two stationarity methods is €4, with the median being the same. The difference in maximum and minimum value is smaller and so is the standard deviation. Overall, contrary to what can be expected, the linear ADF stationarity method does not seem to be worse than the logarithmic ADF. Actually, it could be argued that this is more viable due to the low volatility of portfolio spread.

Table 9 - Several statistical parameters for the final portfolio values consisting of linearly stationary pairs.

Mean	999
Median	1000
Max	1213
Min	916
Std	14
Data points	1857

7 Results – Parameter optimization

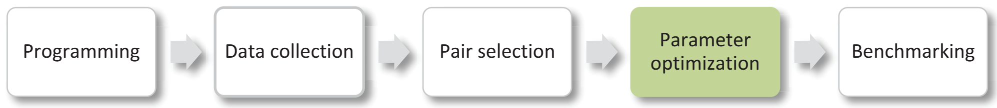

Figure 14 - Work flow chart

In this section, the different parameters that are part of the simulation program are examined and optimized. They are the risk aversion, trading period, parameters estimation period, stationarity period, and transaction frequency. Additionally, we test for two other parameters that are not part of the simulation program per se, but could have equal effect on final portfolio values. These are the level of stationarity (i.e. how statistically certain is it that the random walk hypothesis is false), and correlation. Correlation was originally an independent selection criterion in the previous section, but although it was disqualified, there is still the possibility that correlation has some effect among the pairs selection through stationarity.

7.1 Risk aversion, γ

Although risk-aversion is often occurring in the values for α and β (see Appendix A), we found that the $\frac{1}{1-\gamma}$ fraction component of the weighting function had the largest impact on the optimal weighting of *h*. Figures 15 and 16 show two trading scenarios where, cateris paribus, the risk-aversion is varied. N.B. that the curve has exactly the same shape; the only difference between the two plots is the scaling on the y-axis. Thus, since γ does not affect our comparison, we set it to -3.

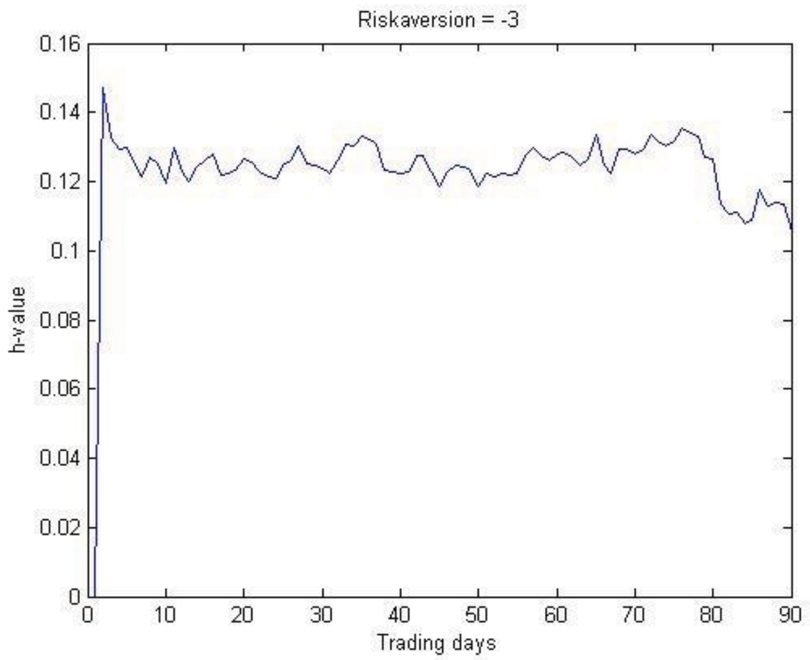

Figure 15 - The h-value throughout a trading period with risk aversion -3.

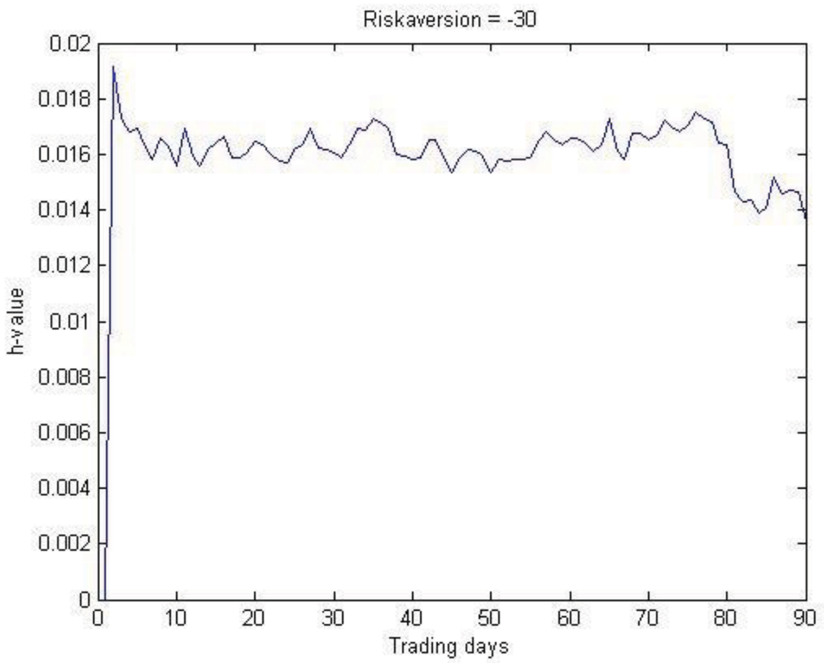

Figure 16 – The h-value throughout a trading period with risk aversion -30.

7.2 Time periods

In this section, we optimize the values for the four time variables as described in part 3.3.4. A short recapitulation could be in place here. The first one is the trading period; how many days the simulation is run, and therefore, towards which day we are optimizing the portfolio value. The second is the parameter estimation period, which is the amount of days we look back to calculate the 6 parameters used in the weighting function. The third one is the stationarity period (T_C) for the spread. This parameter is not part of the trading simulation but rather selection process per se, but since it proved to be of importance in the first study, and since stationarity is assumed in the theory of this stochastic control method we decided to include it here as well. Here, we check how far back in time we look to determine whether the stock pair is stationary or not.

7.2.1 Trading period, T

While higher means and max values occur with increased amount of trading days, this change is not proportional to the day count (There is only an increase 9% of the max and 0.07% in the mean upon doubling the amount of trading days from 60 to 125 for instance). The median remains the roughly same, and the standard deviation fluctuates (table 10).

Table 10 – T = 60,125,251. T_C = 251, N = 125

Mean	1002	Mean	1003	Mean	1003
Median	1001	Median	1000	Median	1000
Std	37	Std	46	Std	43
Max	1542	Max	1677	Max	1760
Min	633	Min	666	Min	723
Data points	1236	Data points	1236	Data points	1217

7.2.2 Parameter estimation period, N

Analogous to the trading period optimization study, increase in parameters causes a steady increase in the mean and the maximum values, while keeping the median steady at 1000 (table 11). The standard deviation increases as well, although not as much as the maximum value (percentage-wise). And further, increasing the parameter estimation period brings up the minimum value significantly, although this difference is only observed in the increase from 60 to 251, and not in 251 to 500.

Table 11 – N = 60, 125,251. T_C = 251, T = 125

Mean	1000	Mean	1002	Mean	1003
Median	1000	Median	1000	Median	1000
Std	19	Std	27	Std	38
Max	1161	Max	1541	Max	1880
Min	588	Min	855	Min	828
Data points	1517	Data points	1517	Data points	1517

7.2.3 Stationarity period, T_c

A longer stationarity period however, does significantly reduce the standard deviation and markedly increases the minimum values, without affecting neither the mean nor the maximum values to a high degree (table 12). It is interesting to note that the maximum values are identical for the 251 and 500 periods, and the means are practically identical as well. Also, the spread for more pairs seem to be stationary in the long term, as compared to the short term, not entirely equivalent of our theory (1217 versus 1517).

Table 12 – T_C = 125, 251, 500. N = 125, T = 125

Mean	1003	Mean	1002	Mean	1002
Median	1000	Median	1000	Median	1000
Std	46	Std	33	Std	27
Max	1677	Max	1541	Max	1541
Min	666	Min	721	Min	855
Data points	1236	Data points	1217	Data points	1517

7.2.4 Transaction frequency, Δt

When decreasing the transaction frequency, we see a decline in mean and max values, although there is an increase in the minimum values and a decrease for the standard deviation (table 13). The decrease in standard deviation is easily explained by fewer trading days.

Table 13 – Δt = 1, 2 and 5. T_C = 251, N and T = 125.

Mean	1004	Mean	1000	Mean	1001
Median	1000	Median	1000	Median	1000
Std	34	Std	26	Std	13
Max	1470	Max	1168	Max	1108
Min	696	Min	827	Min	877
Data points	1236	Data points	1236	Data points	1236

7.3 Level of stationarity

The t-values in the ADF test were also assayed to verify whether more statistically significant stationarity had any impact on the final value of the portfolio. While stationary pairs with low significance have high ending portfolio values as well as low, the ones with high stationarity statistic do not drop as low (figure 17). (Minimum value ~930 for cointegrated pairs with t-values below -6, compared to around 800 for the rest).

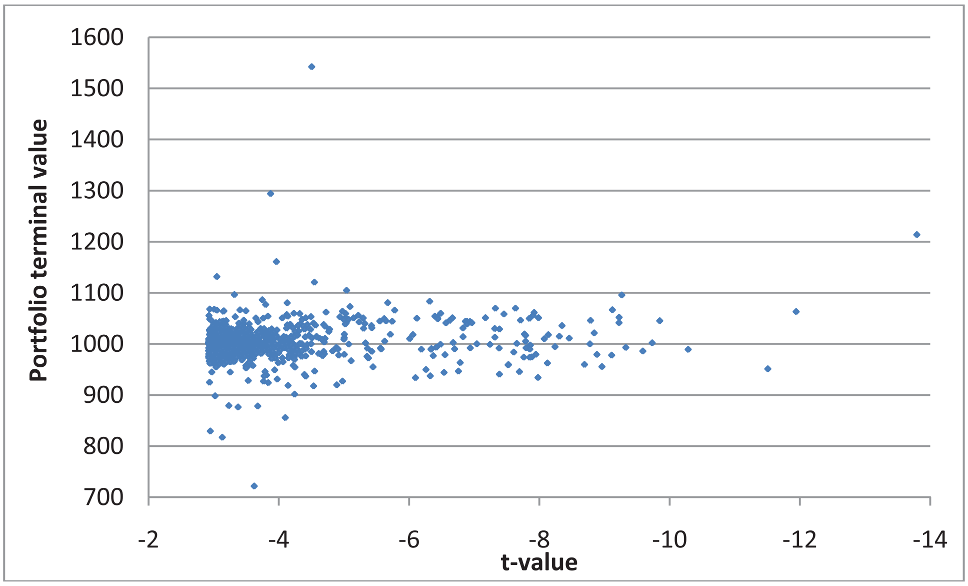

Figure 17 – Terminal portfolio values scattered after the t-value of the stationarity test of their respective stock pairs.

7.4 Correlation

Our hypothesis was that sub-perfect correlation was optimal in this trading strategy, since it would give rise to more profitable trading opportunities providing that the stationarity criterion is fulfilled. Upon validation however, there is no data to back up this hypothesis. No difference was observed in choosing very high correlation, none or even negative. However, stock pairs with correlation coefficients below -0.35 were not observed, indicating they were not stationary (figure 18). The spread towards the high end correlation is just as large as observed earlier where stationarity was not tested. Therefore, we chose not to continue using correlation as a selection criterion, since inadequate pairs are automatically filtered out in the stationarity screen.

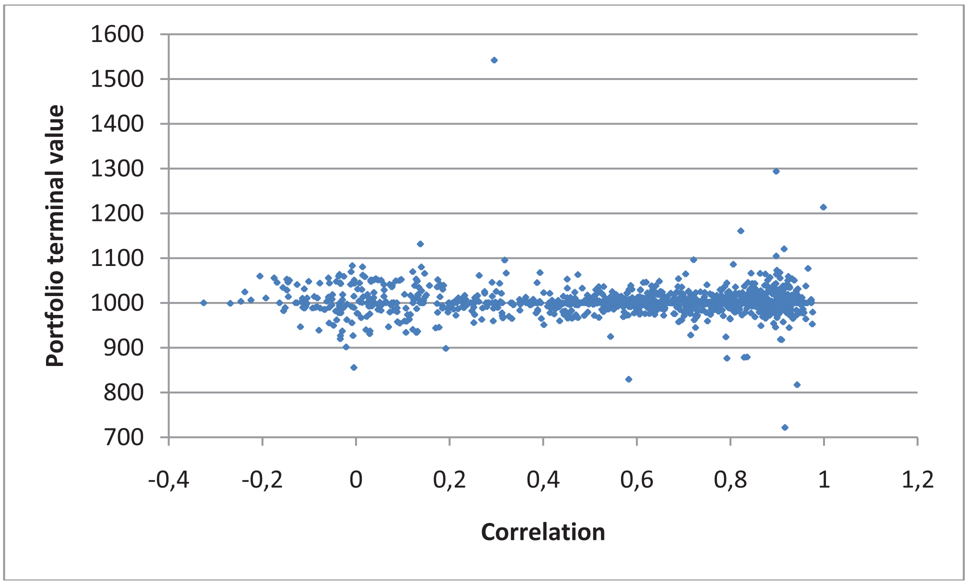

Figure 18 - Terminal portfolio values scattered after the correlation value for stock pairs with stationary spread.

7.5 Bringing the variables together

7.5.1 Trading period

Because of the highest mean, max and first and foremost min values, we choose the longest time period (251 trading days, the equivalent of one year) as the most suitable to run.

7.5.2 Parameter optimization period

The highest parameter value is chosen as well, for the same purpose of the trading period (the data showed a more profitable trend towards the high number, 251)

7.5.3 Stationarity period

Stationarity period is set to 500 days because of the high minimum value, low standard deviation, a wider array of pairs. We choose not to set it any further because extending this period for over than two years decreases the amount of pairs significantly.

7.5.4 Transaction frequency

Given the nature of parameter calculation and the entire strategy outline, we choose to trade every day, despite the transaction costs.

7.5.5 Level of stationarity

Because of lack of overtly poor portfolio developments for stock pairs with higher negative t-values in the ADF stationarity screen, we choose to study pairs that have high stationarity significance. Following the spread as shown in figure 17, we chose to set the cut-off value at -6. This value is almost two times lower than the MacKinnons 1% percentile (-3.50), indicating a very high significance of stationarity.

8 Results – Benchmarking

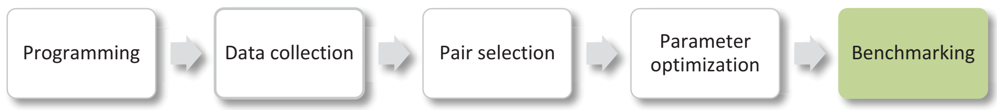

Figure 19 - Work flow chart

In this section we compare the results of our optimization with one of the first and basic pair trading strategies described in section 3.5.

8.1 Optimal parameters

Table 14 summarizes the values to which the different parameters were set in the previous section.

Table 14 – The values of parameters after the independent singular optimization studies.

Level of stationarity	-6
Riskaversion	-3
Trading period	251
Parameters	251
Stationarity period	500
Transaction frequency	1
Correlation	All

Table 15 summarizes the evaluation of the optimization. The amount of pairs are from 19 900 to 73. The corresponding measurements values for these 73 pairs are markedly above the control in almost all aspects. The mean and median are €39 and €24 above the respective values in the negative control. The spread between max and min value was narrower even though the standard deviation was higher. The min value was much higher. The parameter setting this simulation apart from other simulations is the mean value. However, even upon parameter optimization, the annual return rate would just be 2.4%.

Table 15 - Several statistical parameters for the final portfolio values with optimized parameters.

Mean	1024
Median	1008
Std	104
Max	1830
Min	919
Data points	73

8.2 The basic pairs trading strategy

Finally, we compare the stochastic control in pairs trading to the standard deviation strategy[55].

The alternative strategy offers much higher revenue that the SCA (table 16). The mean value shows an average annual return rate of 11.5%, the equivalent of the actual return of relative value strategies used by hedge funds. There were two pairs that deviated significantly from the rest. One increased its portfolio value by 26 times and the other lost €1000 (figure 20). The median value in this simulation is a bit misleading because 431 pair portfolios have an end value above €1000 while only 114 dropped below €1000. This means that 691 pairs with stationary spread were not traded at all.

Table 16 – The result of the portfolio development simulation based on the basic pairs trading strategy. The risk level is the equivalent of the weighting function in the SCA; in this case 80% of the portfolio is invested in stocks at all times. Entry trigger is the standard deviation amount from the mean required to trigger the program to open a position. Analogously, the exit trigger is the amount to which the spread must decrease in order to close the sale.

Mean	1155	Risk level	0.8
Median	1000	Entry trigger	2
Std	965	Exit trigger	1
Max	26233	Days	251
Min	0		
Data points	1236		

[55] Gatev

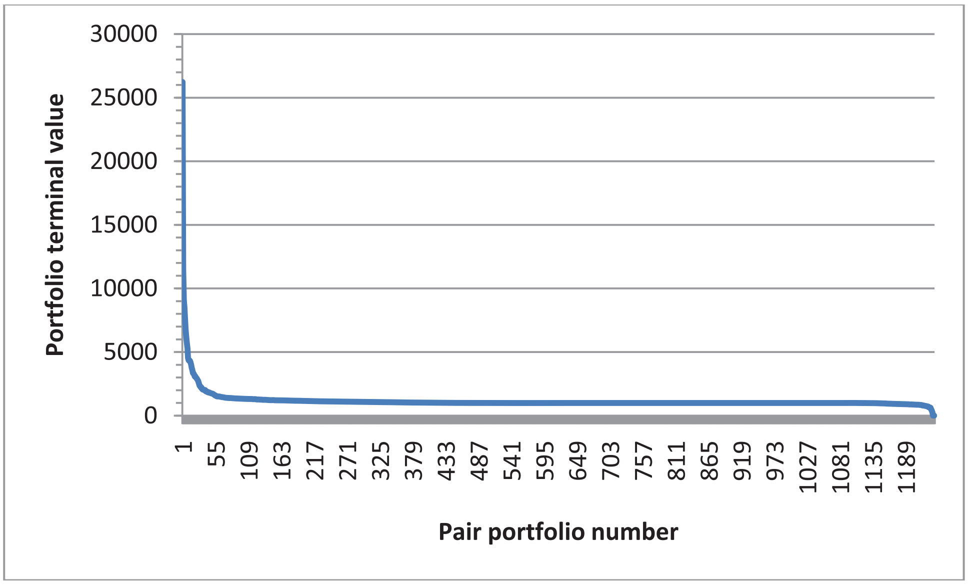

Figure 20 –The terminal portfolio values for the alternative trading strategy.

9 Discussion and contributions

9.1 Selection criteria screen

In the selection criteria screen, ADF stationarity was the only test that showed an improvement of average value to warrant an extension of the study. As we discuss in section 9.2, the ADF stationarity test is the more accurate test of the two.

In the model, there was nothing said about correlation and in line with this, different values for correlation did not affect the terminal portfolio values. A form of spread was actually observed, but we presume that this is due to an inherent characteristic of correlated pairs in relation to stationarity (i.e. high correlated pairs are more probable to have stationary spread, but the reverse is not true), rather than correlation being a suitable selection criteria per se.

Sorting pairs after industries did not produce profit. Likely, this is because the qualitative selection method does not satisfy the requirements of a quantitative model and the strictly defined statistical parameters that come thereof.

9.2 Type and level of stationarity

The fact that linear stationarity was as good selection criteria as the logarithmic came unexpectedly. This implies that perhaps logarithmic stationarity is not a prerequisite. It is possible that just some stability is required, where the spread does not diverge or follow a random-walk pattern too much (i.e. ϕ being close to 1).

The simpler linear regression model did not deliver pairs that gave results as good as the ADF test. One reason being, statistically the linear regression model tests the wrong property. We are able to screen out stocks whose spread are not stationary quite easily, but this does not mean that all the other pairs are. In particular, since logarithmic stationarity did not necessarily mean better mean values compared to the linear stationarity screen, including pairs with spread that potentially may not be stationary is devastating to the average portfolio value.

However, the t-value finding is much more logical. As described in the theoretical framework, we investigate whether the spread follows a random walk pattern in which case it has a unit root. Having a unit root implies that $\phi = 1$ in the equation

$$y_t = \phi y_{t-1} + \varepsilon_t$$

and H_0 is true, otherwise H_0 is false and H_1 is true. However, we know from statistics that just because H_0 is false does not mean that H_1 is necessarily true.

Rather, since H_1 only states that ϕ is smaller than one, we can view the t-value as an indicator of how much smaller than one ϕ is. A high negative t-value means that there is very high certainty that H_0 is false and therefore, ϕ differs greatly from 1, meaning changes in y_t is proportionately more dependent on the random term $\varepsilon \in N(0, \sigma^2)$, in turn implying that the stock pair spreads are more stationary (i.e. stable). And in line with the modeling of the spread, a stable stationarity pattern is more preferable for parameter and weighting function calculations than a more fluctuating one.

9.3 Parameter optimization

In this particular pairs trading strategy it should be noted what generally is considered to be a very low risk-aversion in terms of the absolute value of γ, still does not cause large effects. Clearly, γ has less of an impact in this particular pairs trading strategy than for investment strategies in general, where γ can even determine what stocks to invest in. The more sudden changes in portfolio value appears rather to come from sudden spread divergence rather than changes in h. This forces the investor to make a tough trade-off when setting the risk aversion. Should he or she choose to be risk-averse, then the revenue will be very small, except from the possible leap in the spread that could occur. If he or she goes the other way, there is a risk of high loss should the weighting function to be oriented properly.

The trading period can almost be compared to the investment strategy of a pension fund. When the ending time is far away, more volatile investments can be afforded, but as the ending time comes closer, investments become more and more conservative. In the case of short trading periods, we can explain the lower profit by there not being sufficient time to even out this imbalance, and compensate for the volatile investments that did not become profitable. During a further time period, this discrepancy is better avoided.

The better results obtained from the higher parameter estimation period comes is well in line with the mathematical model. The more values we present, the more probable it becomes that the stocks follow a log-normal distribution. When this is true, the value of each parameter becomes more correctly estimated leading to a better value for the weighting function.

We observed a higher amount of pairs that were stationary when the testing period was two rather than just one year. Although we state in the theory that real world stocks cannot be expected to retain this property for an extended period of time, this proves not to be true in all cases. A reasonable explanation is that during one year the stocks will drift apart asynchronously, but during the second year, their return to a former state makes the spread come closer towards the mean value again. Also, the price data was simulated after the stationary logarithmic spread, to fit the optimization problem. In the Mudchanatongsuk et.al preliminary test study, the simulation was run for 1500 trading days (which approximately equals to 6 years) quadrupling the initial value. This is another explanation to why modeling the parameters and checking for stationarity for the longest time periods yielded the best mean results.

Increasing the trading intervals to two or five days with the goal of decreasing transaction costs was not viable when disregarding transaction costs. Likely, this is because by the second of third day, the equilibrium has shifted to far to make a satisfying prediction and evaluating the weight thereafter. If the algorithm for the parameters in Appendix B are not tweaked, trading should be conducted each consecutive day.

9.4 The stochastic control approach

Practically all mathematical models are a simplification of the real world, in one way, or the other. Firstly, the behavior of the stocks is assumed to follow a predefined pattern. Stock B is assumed to follow a geometric Brownian motion that is not affected by major environmental factors. Instead, the changes in this stock are affected by drift and volatility that are estimated through the maximum likelihood parameter value estimators. And stock B is assumed to be log-normally distributed.

Furthermore the spread is defined as the logarithmic difference between stocks A and B. It is assumed to follow an Ornstein-Uhlenbeck process with a drift, long-term equilibrium, rate of reversion and volatility. Even if making the assumption of stock B being log-normally distributed (a property many simulations seems to ascribe stocks), the same cannot be said about the difference. Moreover, it is dubious why the spread is modeled prior to stock A. This seems like an unnecessary oversimplification of the actual situation, even if makes solving the optimization problem easier.

In the next stage, the behavior of stock A is derived through Ito's lemma. This causes the stock that is in reality should not be more complicated than stock B to have a much more intricate function. This creates an imbalance, making stock A dependent on stock B, which is not necessarily the case. We do however point out that the stocks of different companies in reality can be dependent of each other, but rarely in the one way dependent relationship as in this case of A to B.

Upon formulation of the equation describing the change of the portfolio value, a major restriction is made, fixing the weighting function of A to the weighting function of B but with the opposite sign. It is logical to equalize the weighting functions to keep the strategy market neutral. But we need to point out that this it is a strong constraint regarding the optimization problem formed constraining the optimization problem to not give the best values.

Additionally, it must be emphasized that the optimization problem is directed towards the terminal value of the portfolio. With this in mind, it becomes a logical requirement to allow the weighting function complete freedom, because there is a direct causal relationship between optimizing the weighting function and optimizing the final portfolio value. However in the model, firstly there is the supposed restriction of $|h(t)| \leq 1$, because as is said before, some markets do not allow betting on more than what can be afforded to lose. But more importantly, there is also the model shortcoming of rendering the weighting function infinite if the price for stock B does not change, as described in 6.1. In these two cases, we are forced to change h (in the second case we set the weighting function value to zero). But because of this, the entire optimization problem becomes unbalanced (since we deliberately change the value of h). And changing the intermittent values makes the whole process sub-optimized. Although probably not among the largest contribute factors, we do believe that this still negatively affect the performance of the portfolios.

The rest of the problem is solved without further economic assumptions leading to mathematical simplifications, but limiting the behavior of the two stocks and the weighting function for them makes this strategy applicable only in specific situations. As we showed in the results section, such pairs can be found, but only rarely. Additionally the many layers of models can potentially cause error amplification throughout the simulation.

In Mudchanatongsuk’s study, the model is applied on simulated data, which is fitted to the model. But as we showed, there are rarely over 10% of all initial pairs with a spread that behaves in a statistically significant and logarithmically stationary pattern. Presumably an even smaller fraction of these can be assumed to be log-normally distributed with the mean values and standard deviations as described by the parameters in the paper. This creates a mis-weighting in the portfolio, because the parameters and the weighting function are calculated on the basis of a distribution that does not exist.

9.5 Benchmark with basic pairs trading strategy

The outcome of the benchmark was expected but not the magnitude of it. The SCA was expected to perform better and the reasons behind this have already been discussed. It is however interesting to see that a basic pairs trading strategy with straightforward trading rules with common values still can deliver decent returns. As discussed earlier, widely used arbitrage strategies are chasing the same profit opportunities and therefore should have a harder time to exploit opportunities.

Our results shows that pair trading with straightforward trading rules based on statistical properties of the spread works as long as the stock pairs fulfill stationarity. The drawback though is that basic pairs trading strategies lack predictive ability on investment horizon because no economic models of price patterns, spread behavior etc. are used.[56]

Both strategies are consistent with market neutrality since they gave positive mean return on market data from the period of the recent financial crisis.

9.6 Improvements

Two general improvement strategies can be applied here. First, in order to achieve the same revenue as the basic pairs trading strategy, the model will have to be of broader kind. Some of the behavioral restrictions on the stocks and the spread will have to be lifted, and the causal relationship between stock A and B through the spread will have to be redefined. Also, we strongly suggest removing the weighting equality constraint when solving the problem.

Alternatively, the stocks will need to be evaluated more thoroughly. The presented performance (38% compounding return rate) of the SCA in Mudchanatongsuk et al. is fairly impressive. A questionable practice is the generation of artificial price data modeled the same way as the stochastic control problem is. The simulated data is based on the log-normal spread where the mean and the variance are described through the six parameters that are estimated at a later time point. This results in an optimal weighting based on the very same price dynamics that are used for simulating price series which not surprisingly will give good returns. Thus, to acquire the same annual return rate as in the simulated run on imaginary data, one would have verify that the values of stock B and the spread follows a log-normal distribution with the same interparametral distribution as described by Mudchanatongsuk.

An interesting observation by Papadakis & Wysocki is that pair trades triggered after accounting events such as earnings announcements are less profitable.[57] This will have an impact on the degree of automation of the strategy. Some hedge funds plug out their automatic trading algorithms at times around news releases to avoid the strong irregular price movements.

[56] Do (2006)
[57] Papadakis & Wysocki (2007)

We show in this study that theoretical models do not necessarily describe the real world in such a good way that a strategy formulated on model assumptions will be profitable. The profit observed was judged to be too small to view the strategy as successful. However, when evaluating the controlling basic pairs trading strategy, we found that a simple algorithm can achieve comparable profits as with what investors would expect from this type of strategy.

10 Conclusions

In this study, we have used a stochastic control method to optimize the final portfolio value in the pairs trading investment strategy. We have obtained data for 200 of the most liquid stocks, resulting in possible 19 900 pairs. The pairs were screened for properties that would make them suited for use in this particular stochastic control algorithm. The model assumed a stationary logarithmic spread and had several parameters such as the amount of trading days, risk aversion and parameters estimation period. We modeled the effect of each parameter and screening strategy individually, and then made a simulation where all pairs where put to their optimal value.

The final simulation gave higher values in the mean, median, max and min values as compared to the controls. However, the mean value was below the much simpler basic pairs trading strategy. While we were able to find a method for choosing pairs, resulting in higher values than the intrinsic controls, the non-optimized basic strategy outperformed the optimized form of the stochastic one. To possibly counter this effect, we suggested two different strategies that can be used to improve the SCA; lifting some of the restrictions or subjecting the stocks and their spread to more rigorous studies.

This report is an addition to the pool of evidence that many academic models are not applicable in real life. Perhaps we have managed to describe in deeper detail why this is the case. We hope that subsequent studies will follow this one, revising the selection criteria and the assumptions made when finding the solution for the optimal weighting function.

As final words it can be said that the market competition is fierce for quantitative strategies, especially since the introduction of electronic trading. The results of this study show that this is a sign of a possible over-engineering phenomenon that exists in the market in competition for investment opportunities.

Bibliography

Published sources

Bock, Michael & Mestel, Roland (2008) “A regime-switching relative value arbitrage rule”, *Operations Research Proceedings*, Springer Berlin Heidelberg, pp. 9-14, August 2009.

Bryman, A. & Bell, E. (2007) *Business Research Methods*, Oxford University Press, 2nd edition.

Chan, Kalok & Cung, Y. Peter & Johnson, Herb (1995) “The Intraday Behavior of Bid-Ask Spreads for NYSE Stocks and CBOE Options”, *The Journal of Financial and Quantitative Analysis*, Vol. 30, No.3, pp. 329-346, Sep. 1995.

Do, Bhinh & Faff, Robert & Hamza, Kais (2006) “A New Approach to Modeling and Estimation for Pairs Trading”, Working Paper, Monash University.

Engle, F.E. & Granger, C.W.J. (1987) “Co-Integration and Error Correction: Representation, Estimation and Testing”, *Econometrica: Journal of the Econometric Society,* 55:2, pp. 251-276.

Fama, E.F. (1969) “Efficient Capital Markets: A Review of Theory and Empirical Work”, *The Journal of Finance*, Vol. 25, No.2, Papers and proceedings of the Twenty-Eighth Annual Meeting of the American Finance Association New York, N.Y. December, 28-30, 1969.

Gatev, Evan G. & Goetzmann, William N. & Rouwenhorst, K. Geert (1999) “Pairs Trading: Performance of a Relative Value Arbitrage Rule”, Working Paper, National Bureau of Economic Research.

Granger, C.W.T. (2004) “Time series analysis, cointegration, applications”, *The American Economic Review*, 94:3, pp. 421-425.

Groβ, J. (2003), *Linear Regression,* Springer.

Hamilton, J. (1994) *Time series analysis*, Princeton University press, USA.

Hull, John C. (2005) *Options, Futures and Other Derivatives*, Prentice Hall, 6th edition.

Ineichen, A (2000) *In Search of Alpha – Investing in Hedge Funds*, UBS Warburg.

Ineichen, A. & Silberstein, K. (2008) *AIMA's Roadmap to Hedge Funds*, Alternative Investment Management Association.

Jacobs, Bruce I. & Levy, Kenneth N. (1993) "Long/Short Equity Investing", *Journal of Portfolio Management,* Vol. 20(1), pp. 52-64.

Johansen, S. (1991) "Estimation and Hypothesis Testing of Cointegration Vectors in Gaussian Vector Autoregressive Models", *Econometrica,* 59:6, pp. 1551-1581.

Kissell, Robert & Glantz, Morton & Malamut, Robert (2003) "A practical framework for estimating transaction costs and developing optimal trading strategies to achieve best execution", *Finance Research Letters 1*, pp. 35-46, 2004.

Kiyoshi, Ito (1951) "On stochastic differential equations", *Memoirs, American Mathematical Society 4*, pp. 1-51.

Lee, Charles M. C. & Mucklow, Belinda & Ready, Mark J. (1993) "Spreads, Depths, and the Impact of Earnings Information: An Intraday Analysis.", *Review of Financial Studies vol. 6*, pp. 345-374, 1993.

MacKinnon, J.G. (1996) "Numerical Distribution Functions for Unit Root and Cointegration Tests", *Journal of Applied Econometrics*, Vol. 11, pp. 601-618.

Mudchanatongsuk, Supakorn & Primbs, James & Wong, Wilfred (2006) "Optimal Pairs Trading: A Stochastic Control Approach", *Proceedings of the American Control Conference*, Seattle, WA, pp. 1035-1039, June 2008.

Mudchanatongsuk, Supakorn & Primbs, James A. & Wong, Wilfred (2008) "Optimal Pairs Trading: A Stochastic Control Approach", *Proceedings of the American Control Conference*, Seattle, WA, pp. 1035-1039, June 2008.

Papadakis, G. & Wysocki, P (2007) *Pairs Trading and Accounting Information*, Working paper, Boston University & MIT.

Purnendu, Nath (2003) *High Frequency Trading with U.S. Treasury Securities: Risks and Rewards for Hedge Funds*, Working paper, London Business School.

Vidyamurthy, Ganapathy (2004) *Pairs trading – Quantitative Methods and Analysis*, Wiley Finance.

Øksendal, Bernt (2003) *Stochastic Differential Equations: An Introduction with Applications*, Springer, 6th edition.

Electronic sources

AMF press announcement: *Short-selling: Ban on unsecured transactions and transparency of short positions in financialsector securities* (2008-09-19), retrieved (2010-04-03) from Autorité des marchés financiers: http://www.amf-france.org/documents/general/8424_1.pdf

ASIC press announcement: Covered short selling not permitted (2008-09-21), retrieved (2010-04-03), from Australian Securities & Investments Commission: http://www.asic.gov.au/asic/asic.nsf/byheadline/08-205+Covered+short+selling+not+permitted?openDocument

BaFin press announcement: *General Decree of 30 March 2009 on the Extension of the Provisions of the General Decrees of the Federal Financial Supervisory Authority (BaFin) of 19 and 21 September 2008* (2009-03-30), retrieved (2010-04-03) from Bundesanstalt für Finanzdienstleistungsaufsicht: http://www.bafin.de/cln_152/nn_720788/SharedDocs/Aufsichtsrecht/EN/Verfuegungen/vf__090330__leerverk__verlaeng__en.html.

FSA press announcement: *FSA statement on short positions in financial stocks* (2008-09-18), retrieved (2010-04-03), from Financial Services Authority: http://www.fsa.gov.uk/pages/Library/Communication/PR/2008/102.shtml

London Stock Exchange website – *Our history* (2010), retrieved (2010-04-15): http://www.londonstockexchange.com/about-the-exchange/company-overview/our-history/our-history.htm

New York Stock Exchange & Euronext: *History* (2010), retrieved (2010-04-15): http://www.nyse.com/about/history/timeline_1980_1999_index.html

SEC press announcement: *SEC Halts Short Selling of Financial Stocks to Protect Investors and Markets* (2008-09-19), retrieved (2010-04-03), from Securities and Exchange Commission: http://www.sec.gov/news/press/2008/2008-211.htm

Appendix A – Companies

1	FRANCE TELECOM	36	PPR
2	A TOUTE VITESSE - ATV	37	EDF
3	TOTAL	38	ATOS ORIGIN
4	ALCATEL-LUCENT	39	APRR
5	AXA	40	TECHNIP
6	VIVENDI	41	LAGARDERE GROUPE
7	SANOFI-AVENTIS	42	THALES
8	GDF SUEZ	43	RHODIA
9	BNP PARIBAS	44	SAFRAN
10	CREDIT AGRICOLE	45	ATARI
11	CARREFOUR	46	VALEO
12	SOCIETE GENERALE	47	CASINO GUICHARD-PERRACHON
13	DANONE	48	DEXIA
14	SAINT GOBAIN	49	ALTRAN TECHNOLOGY
15	EADS (PAR)	50	HAVAS
16	TECHNICOLOR	51	SODEXO
17	LAFARGE	52	ESSILOR INTL.
18	RENAULT	53	PAGESJAUNES
19	L'OREAL	54	DASSAULT SYSTEMES
20	ALSTOM	55	MAUREL ET PROM
21	VINCI (EX SGE)	56	VALLOUREC
22	LVMH	57	SOITEC
23	CAP GEMINI	58	EIFFAGE
24	VEOLIA ENVIRONNEMENT	59	UNIBAIL-RODAMCO
25	PEUGEOT	60	AVENIR TELECOM
26	AIR LIQUIDE	61	CHRISTIAN DIOR
27	BOUYGUES	62	GROUPE EUROTUNNEL
28	STMICROELECTRONICS	63	SCOR SE
29	SCHNEIDER ELECTRIC	64	M6-METROPOLE TV
30	MICHELIN	65	EURO DISNEY SCA
31	ACCOR	66	CNP ASSURANCES
32	AIR FRANCE-KLM	67	ZODIAC AEROSPACE
33	TF1	68	NEOPOST
34	PERNOD-RICARD	69	UBISOFT ENTERTAINMENT
35	PUBLICIS GROUPE	70	CGGVERITAS

71	IMERYS
72	JCDECAUX
73	ILIAD
74	HERMES INTL.
75	GFI INFORMATIQUE
76	HAULOTTE GROUP
77	GENERALE DE SANTE
78	WENDEL
79	INGENICO
80	BOURBON
81	NEXANS
82	TELEPERFORMANCE
83	CARBONE-LORRAINE
84	GROUPE STERIA SCA
85	KLEPIERRE
86	EURAZEO
87	HI MEDIA
88	BIC
89	RODRIGUEZ GROUP
90	GEMINA
91	BULL REGPT
92	REMY COINTREAU
93	CIMENTS FRANCAIS
94	ANOVO
95	FAURECIA
96	NEXITY
97	ARTPRICE.COM
98	NRJ GROUP
99	AREVA CI
100	ALTEN
101	SIEMENS
102	E.ON
103	DAIMLER
104	DEUTSCHE TELEKOM
105	BASF
106	SAP
107	RWE
108	BAYER
109	MÜNCHENER RÜCK
110	COMMERZBANK
111	DEUTSCHE POST
112	BMW
113	CONTINENTAL
114	DEUTSCHE BÖRSE
115	ADIDAS
116	VOLKSWAGEN
117	THYSSENKRUPP
118	MAN
119	HYPO REAL ESTATE BANK
120	METRO
121	DEUTSCHE LUFTHANSA
122	LINDE
123	HENKEL
124	INFINEON TECHNOLOGIES
125	DEPFA BANK
126	TUI
127	MERCK KGAA
128	ALTANA
129	PUMA
130	FRESENIUS MEDICAL CARE
131	CELESIO
132	FRESENIUS PREF.
133	DEUTSCHE POSTBANK
134	LANXESS
135	K + S
136	HOCHTIEF
137	RHEINMETALL
138	BEIERSDORF
139	SALZGITTER
140	HEIDELBERGER DRUCKMASCHINEN

141 HANNOVER RÜCK
142 IVG IMMOBILIEN
143 BILFINGER BERGER
144 FRAPORT
145 STADA ARZNEIMITTEL
146 WINCOR NIXDORF
147 QIAGEN
148 RHÖN-KLINIKUM
149 BB BIOTECH
150 GENERALI DTL.HLDG.
151 SÜDZUCKER
152 SMARTRAC
153 GEA GROUP
154 UNITED INTERNET
155 IKB DEUTSCHE INDUSTRIEBANK
156 MLP
157 DOUGLAS HOLDING
158 PROSIEBEN SAT 1 MEDIA
159 SOLARWORLD
160 MTU AERO ENGINES HLDG.
161 SOFTWARE
162 HUGO BOSS
163 SGL CARBON
164 AAREAL BANK
165 DEUTSCHE EUROSHOP
166 LEONI
167 FREENET
168 PFLEIDERER
169 SINGULUS TECHNOLOGIES
170 KUKA
171 VIVACON
172 FUCHS PETROLUB
173 VOSSLOH
174 GFK
175 KRONES
176 PFEIFFER VACUUM TECH.
177 BAUER
178 BAYWA
179 EADS
180 ELRINGKLINGER
181 GAGFAH
182 GERRESHEIMER
183 SKY DEUTSCHLAND
184 GILDEMEISTER
185 BECHTLE
186 AIXTRON
187 ALLIANZ
188 BALDA
189 DEUTSCHE BANK
190 DRAEGERWERK
191 FIELMANN
192 IDS SCHEER
193 JENOPTIK
194 KONTRON
195 MEDION
196 MORPHOSYS
197 MPC MÜNCHMEYER
198 Q-CELLS
199 RATIONAL
200 QSC

Appendix B – Parameter estimation formulas

The appendix provides an in-depth derivation of all constants and functions used in the simulation.

$$\hat{\sigma} = \sqrt{\frac{\hat{S}^2}{\Delta t}}$$

$$\hat{\mu} = \frac{\widehat{m}}{\Delta t} + \frac{1}{2}\hat{\sigma}^2$$

$$\hat{k} = -\frac{\log(\hat{p})}{\Delta t}$$

$$\hat{\theta} = \frac{\hat{q}}{1-\hat{p}}$$

$$\hat{\eta} = \sqrt{\frac{2\hat{k}\hat{V}^2}{1-\hat{p}^2}}$$

$$\hat{\rho} = \frac{\hat{k}\hat{C}\hat{V}\hat{S}}{\hat{\eta}\hat{\sigma}(1-\hat{p})}$$

Where

$$\widehat{m} = \frac{S(N) - S(0)}{N}$$

$$\hat{S}^2 = \frac{\sum_{t=0}^{N-1}(S(t+1) - S(t))^2 - 2\widehat{m}\big(S(N) - S(0)\big) + N\widehat{m}^2}{N}$$

$$\hat{p} = \frac{(\sum_{t=0}^{N-1} X(t)\Delta s_t - \widehat{m}\sum_{t=0}^{N-1} X(t))(N\sum_{t=0}^{N-1} X(t+1)\,\Delta s_t) - \widehat{m}(X(N) - X(0) + \sum_{t=0}^{N-1} X(t))(N\sum_{t=0}^{N-1} X(t)\,\Delta s_t)}{(\sum_{t=0}^{N-1} X(t)\,\Delta s_t - 2\widehat{m}\sum_{t=0}^{N-1} X(t))(N\sum_{t=0}^{N-1} X(t)\Delta s_t) - N^2\hat{S}^2(\sum_{t=0}^{N-1} X(t)^2) + (\sum_{t=0}^{N-1} X(t))^2(\sum_{t=0}^{N-1}(\Delta s_t)^2)} -$$

$$\frac{N^2\hat{S}^2(\sum_{t=0}^{N-1} X(t+1)X(t)) + (X(N) - X(0) + \sum_{t=0}^{N-1} X(t))(\sum_{t=0}^{N-1} X(t))(\sum_{t=0}^{N-1}(\Delta s_t)^2)}{(\sum_{t=0}^{N-1} X(t)\,\Delta s_t - 2\widehat{m}\sum_{t=0}^{N-1} X(t))(N\sum_{t=0}^{N-1} X(t)\Delta s_t) - N^2\hat{S}^2(\sum_{t=0}^{N-1} X(t)^2) + (\sum_{t=0}^{N-1} X(t))^2(\sum_{t=0}^{N-1}(\Delta s_t)^2)}$$

$$\Delta s_t = S(t+1) - S(t)$$

$$\hat{q} = \frac{X(N) - X(0) + \sum_{t=0}^{N-1} X(t) - \hat{p}\sum_{t=0}^{N-1} X(t)}{N}$$

$$\hat{V}^2 = \frac{1}{N}\left[X^2(N) - X^2(0) + (1+\hat{p})\sum_{t=0}^{N-1} X^2(t) - 2\hat{p}\sum_{t=0}^{N-1} X(t)X(t+1) - N\hat{q}\right]$$

$$\hat{C} = \frac{1}{N\hat{V}\hat{S}}\left[\sum_{t=0}^{N-1} X(t+1)\Delta s_t - \hat{p}\sum_{t=0}^{N-1} X(t)\Delta s_t - \widehat{m}\big(X(N) - X(0)\big) - \widehat{m}(1 - \hat{p})\sum_{t=0}^{N-1} X(t)\right]$$

Appendix C – Alpha and beta in weighting h

$$\alpha(t) = \frac{k(1-\sqrt{1-\gamma})}{2\eta^2} \cdot \left\{ 1 + \frac{2\sqrt{1-\gamma}}{1-\sqrt{1-\gamma} - (1+\sqrt{1-\gamma})e^{\frac{2k(T-t)}{\sqrt{1-\gamma}}}} \right\}$$

$$\beta(t) = \frac{1}{2\eta^2 \left[\left(1-\sqrt{1-\gamma}\right) - (1+\sqrt{1-\gamma})e^{\frac{2k(T-t)}{\sqrt{1-\gamma}}} \right]} \cdot$$

$$\left\{ \gamma(\sqrt{1-\gamma}(\eta^2 + 2\rho\sigma\eta) \left[1 - e^{\frac{2k(T-t)}{\sqrt{1-\gamma}}} \right]^2 - \gamma(\eta^2 + 2\rho\sigma\eta + 2k\theta) \left[1 - e^{\frac{2k(T-t)}{\sqrt{1-\gamma}}} \right] \right) \right\}$$