UPPSALA
UNIVERSITET

U.U.D.M. Project Report 2020:47

Mathematical aspects of Pairs Trading

Georgios Orfanidis

Examensarbete i matematik, 30 hp
Handledare: Maciej Klimek
Examinator: Erik Ekström
September 2020

Department of Mathematics
Uppsala University

To my family

Abstract

In the current low interest-rate and highly-regulated environment, investing capital efficiently is one of the most important challenges, which most of the financial institutions face nowadays. For that reason, hedge funds, banks, as well retail investors seek for quantitative methods in order to enhance the probability for profit. In this study, a quantitative strategy called 'Pairs Trading' is described and its mathematical aspects are reviewed. Additionally, focus is placed on the performance of the trading strategy. Furthermore, among the traditional quantitative approach of co-integration, another one, based on the Kalman filters will be presented.

Going through the study, the first chapter refers to the introduction. The following chapter deals with the definition and the historical development of statistical arbitrage. Furthermore, a literature review for the different methods of statistical arbitrage is presented. The third chapter addresses the general idea about economics and then the mathematical aspect of co-integration method is examined. The following chapter deals with the mathematical aspect in more detail of the co-integration approach and also with the implementation of it, based on two artificial securities. In the next chapter, the Kalman filter parameter estimation approach is coming along with its derivation. Chapter 6 refers to the empirical analysis, based on two ETFs (Exchanged Traded Funds). The last chapter takes the opportunity of the 2020 crisis and its high volatility, in order to investigate whether a portfolio of pairs is beta neutral, as this is the main building block of the method. In the last chapter, there is some general discussion along with future recommendations.

The main conclusion is that pairs trading seems to have a very promising performance, either the market remains calm or it fluctuates in a more aggressive manner. What is more, throughout our empirical analysis, the beta neutral characteristic of this strategy is being confirmed. Additionally in the current study, transactions costs have not been taken into account.

Keywords: pairs trading, statistics, co-integration, quantitative methods, Kalman filter

Acknowledgements

First of all, I would like to thank Professor Marcej Klimek for his invaluable help in conducting this study. Furthermore, I would like to thank MEAG company and especially my former colleagues from portfolio management department for their helpful comments and feedback, as well as for feeding me with actual data.

Lastly, I thank my family for their endless support throughout my studies.

Contents

List of Figures

Chapter 1

Introduction

Undoubtedly financial markets nowadays have significantly changed. We have witnessed that physical markets have been replaced by virtual. Going back in history, physical markets have been existing since the ancient world. There are many reports in the literature, mentioning that one of the oldest markets was located in ancient Greece and more precisely in Athens. The Athenians used to gather in "agora", which in modern Greek means marketplace, trading olive oil, honey as well as other agricultural products. However, the situation has changed rapidly nowadays. Virtual markets have developed, the products have become very complex and different kind of strategies are used. However, what matters more nowadays is the speed.

In today's modern era, the issue about finding out methods and building models in order to beat the market is a controversial one. The competition between financial institutions and retail investors is enormous. People try to explore ways to predict or to find patterns in a turbulent structure of financial time series. However, many believe that using quantitative models, might bring you ahead of all the others and to make you succeed in having a positive expected excess return over the risk free rate.

In this study, one of the most well-known quantitative strategies, called statistical arbitrage will be described. Moreover, a special case of statistical arbitrage, called pairs trading will be used for our empirical analysis. The main idea of pairs trading is very trivial. As a first step, two securities usually from the same sector, although this is not necessary, are being selected. Then, if there exists a relationship between them, which has a long-run mean reverting behaviour, a predetermined rule can be established and the pair with a specific price relationship can be traded. It is quite common that from time to time, the relationship moves away from its mean, because of changes in supply and demand or because of important economic news. In this scenario, we have statistical evidence that the relationship of two securities is going to converge again to its mean and therefore we can take advantage of a temporary price divergence. So, under these circumstances, the under-performing security is being bought and the out-performing security is being sold. At the time it converges to its mean, the positions can be closed and a pair trade can be made.

Obviously it is crucial to determine whether a pair of securities follows a mean reverting behaviour or not. In order to test it, the co-integration approach is highly used nowadays. This is also the approach that will be described throughout this study. According to this framework, two securities are co-integrated if there is a linear relationship between them, which has a constant mean and variance. In other words, we are looking for a linear relationship of them, which is stationary over time. Mathematically speaking, combining the logarithmic prices of two assets by using the OLS regression, as it is suggested by Vidyamurthy (2004) [1], we get the following relationship:

$$\log(Y_t) = \alpha + \beta \log(X_t) + \epsilon_t, \tag{1.1}$$

where ϵ_t is a stationary process with mean equal to zero. Also, ϵ_t is referred as the residual term and β as hedging ratio. Having estimated the coefficients α and β, the final portfolio, also called as spread portfolio, $\log(Y_t) - \beta \log(X_t)$ will oscillate around a statistical equilibrium. Based on this relationship, each time the residual term is small enough according to a trading rule, we go long one unit of Y_t and short β units of X_t. On the other hand, when the residual term is getting large, we long β units of X_t and short one unit of Y_t.

The co-integration approach will be analysed in more details on the following chapters. What is more, in this study, another approach of estimating the hedging ratio β, will be described. The method is called Kalman filter and it is an algorithm, which is used to estimate states based on linear dynamical systems in state space format. At the end of this study, we are going to test the performance of the co-integration method based on a real pair of securities.

Chapter 2

Statistical Arbitrage

Chapter 2 focuses on a general literature review of statistical arbitrage. The chapter starts with a historical view and then tries to capture a way to define statistical arbitrage. At the end, a brief description of the different approaches to statistical arbitrage is presented.

2.1 History

Pair trading was introduced around 1980 and it was pioneered by a quant named Gerry Bamberger. At that time, Bamberger had been working for Morgan Stanley when he came up with the idea of making profit by hedging his position with assets from the same sector. Later on, his colleagues, who were a group of mathematicians, physicists and computer scientists, lead by Nunzio Tartaglia, developed further that idea and eventually constructed algorithms, which were able to execute trades automatically. The main idea behind their attempt was to find two securities which move relatively together over time and trade them according to a specific rule when an anomaly of their relationship was being noticed, with the expectation of its correction in the near future. The following years, this strategy had become an object of research in the academic world, while in the meantime, hedge funds and other financial institutions extensively used it, in order to make profit.

Although, the concept of statistical arbitrage started from Gerry Bamberg, statistical arbitrage came actually to the fore as a trading strategy of a hedge fund named Long Term Capital Management (LTCM). At that moment Nobel Prize winners Scholes and Merton used to both work there. The company focused on fixed income securities, where the complex strategies which were developed, proved very successful until around 1998. At that time, a financial crisis hit countries from east Asia and Russia. As a result of this, strategies of LTCM started generating losses which placed a big risk not only on domestic markets but on global as well. In order to avoid a global financial crisis and a deep recession, the Federal reserve Bank of New York decided to organise a bailout program. Since then, Statistical Arbitrage development continued at a very fast pace, covering also other asset classes like equities, commodities or even cryptocurrencies. Furthermore, nowadays Statistical Arbitrage took advantage of the technological revolution we are experiencing and it is used very frequently in high frequency trading by banks and hedge funds across the world. Furthermore, in the research field, there are a plethora of academic and working papers trying to go deep and to extend the topic of statistical arbitrage.

2.2 Definition

The concept of arbitrage is fundamental in literature and has been used in tens of thousands analyses so far. Although statistical arbitrage has been reviewed by not only academics but also by practitioners, it is quite hard to find exactly one definition. In the literature we can find a relatively large number of them. A study by Avellaneda and Lee [16] showed that the term statistical arbitrage includes different kind of strategies characterized by systematic trading signals, market neutral trades and statistical methods. Montana [17] has defined statistical arbitrage as an invest-

ment strategy that exploits patterns detected in financial data streams. According to Burgess [18], statistical arbitrage is described as a generalization of a traditional arbitrage, where mispricing is statistically determined through replicating strategies. Thomaidis and Kondakis [19] concluded statistical arbitrage as an attempt to profit from pricing discrepancies that appear in a group of assets. Do, Faff and Hamza [20] claimed that statistical arbitrage is an equity trading strategy that employs time series methods to identify relative mispricings between stocks. By using derivatives, Zapart [22] defined statistical arbitrage as an investment opportunity when perfect hedging is not possible. As it can be easily observed from above a variety of definitions on the aspect of statistical arbitrage has been given so far.
Before we present a formal mathematical definition, it is crucial to find out what arbitrage is. According to Björk [12], the definition is given below. Furthermore, it is interesting also to see how a self-financing portfolio has been defined, since the concept of arbitrage is based on a self financing portfolio.

Definition 2.2.1 (Self-financing) *A self financing portfolio h, is a portfolio such that its market value V at time t, equals the purchase value of the new portfolio h at time $t + \Delta t$. Therefore the following conditions should hold:*

$$V_t^h = V_{t+\Delta t}^h$$

Definition 2.2.2 (Arbitrage) *An arbitrage opportunity is a self financing portfolio h with zero value at time $t = 0$, which has a positive value at time $T > 0$ with a positive probability. Therefore the following conditions should hold:*

i. $V_0^h = 0$
ii. $P(V_T^h \geq 0) = 1$
iii. $P(V_T^h > 0) > 0$

Definition 2.2.3 (Statistical Arbitrage) *In a popular study by Jarrow et al. [23], statistical arbitrage is defined as a self financing trading strategy with zero initial cost, which has cumulative discounted trading profits $V(t)$ such that:*

i. $V(0) = 0$
ii. $\lim_{t\to\infty} E[V(t)] > 0$
iii. $\lim_{t\to\infty} P(V(t) < 0) = 0$
iv. $\lim_{t\to\infty} Var[\Delta V(t)|\Delta V(t) < 0] = 0$

Therefore, by Jarrow et al. [23] statistical arbitrage in the first axiom requires the cumulative discounted profits to be zero at time $t = 0$. Then, the expected value of the cumulative discounted profit is positive, as time t converges to infinity. Furthermore, according to the third axiom the probability of a loss converges to zero and the last axiom shows that the variance of incremental profits, given that the incremental profits are negative, converges also to zero. Jarrow et al. [23] suggests on the fourth axiom that investors are only concerned about the variance of a potential decline in wealth.

Although, statistical arbitrage has been defined over an infinite time horizon, there is a time point t^* where the probability of a loss is relatively small, $P(V(t^*) < 0) = \epsilon$, with $\epsilon > 0$. On the other hand, as we have already seen, standard arbitrage generates the opportunity of a profit with zero probability of a loss. Therefore, we could conclude that as the t tends to infinity, the statistical arbitrage coincides with standard arbitrage.

2.3 Pairs trading approaches

Since the method of pairs trading was invented, many studies worked on this topic. All of them tried to approach the problem in a different way. The most well known approaches are named co-integration method, time series approach, stochastic control and distance method. Below we will briefly discuss the main points of them. However, we are not going to cover deeply the topic for each of them since the current study is based only on the co-integration method.

2.3.1 Stochastic control

According to Elliot et al. [26], the spread of two financial assets can be modeled as a mean reverting process. The spread, y_t is defined as the difference of two observed asset prices and it is assumed to be driven by a state process and some additional error measurement, which can be considered as Gaussian noise. So, we get:

$$y_k = x_k + M w_k, \tag{2.1}$$

where y_k is the difference of the two financial quantities, x_k represents the state process at time τ_k for $k = 0, 1, 2...$, M a constant value with $M > 0$, and w_k independent and identical random variables, where $w_k \sim \mathcal{N}(0, 1)$.

The state process x_k is assumed to follow the process which can be written as follows:

$$x_{k+1} - x_k = (\alpha - b x_k)\tau_k + \sigma\sqrt{\tau_k}\epsilon_{k+1}, \tag{2.2}$$

where α, b, $\sigma > 0$ and $\epsilon_k \sim \mathcal{N}(0, 1)$ are independent and identical random variables and independent from w_k. The time increment is defined as τ_k with $k = 1, 2....$ The process mean reverts to $\mu = \frac{a}{b}$ with 'strength' b. Clearly, $x_k \sim \mathcal{N}(\mu_k, \sigma_k^2)$. Since the process (2.2), as k tends to infinity, converges to:

$$\lim_{k\to\infty} \mu_k = \frac{\alpha}{b}, \tag{2.3}$$

which means that the speed of the process can be defined as b. Also,

$$\lim_{k\to\infty} \sigma_k^2 = \frac{\sigma^2\tau}{1-(1-b\tau)^2} \tag{2.4}$$

Furthermore, equation (2.2) can be written through the following format:

$$x_{k+1} = A + Bx_k + C\epsilon_k, \tag{2.5}$$

where $A = \alpha\tau \geq 0$, $0 < B = 1 - b\tau < 1$, $C = \sigma\sqrt{\tau}$. One can regard that $x_k \cong X(k\tau)$ and $\{X(t)|t \geq 0\}$.
Moving to continuous time, the equation [2.2] can be written as a continuous process which follows a stochastic differential equation. The SDE is assumed to follow an Ornstein-Uhlenbeck process such that:

$$dx_t = \kappa(\theta - x_t)dt + \sigma dW_t, \tag{2.6}$$

where the $\kappa(\theta - x_t)$ is the drift term that represents the expected instantaneous change in the spread at time t, θ is the long-term equilibrium level to which the spread reverts and $\{W(t)|t \geq 0\}$ is a standard Brownian motion (on some probability space). The rate of reversion is represented by the parameter κ which has to be positive to ensure stability around the equilibrium value. Furthermore, the parameter σ illustrates the volatility.

It is also defined $Y_k = \{y_0, y_1, ..., y_k\}$ as the set of all the information from observing $y_0, y_1, ..., y_k$ and the conditional expectation as:

$$\hat{x_k} = E[x_k|Y_k], \tag{2.7}$$

where $\hat{x_k}$ is the estimate of the hidden process [2.2] through the observed process [2.1].

Using the Ornstein-Uhlenbeck process, the coefficients A, B, C can be estimated by the state space model [5.3]. The trading philosophy is almost the same across the various pairs trading approaches. In this case, a trading signal is triggered as soon as $y_k \geq pE[x_k|Y_{k-1}]$ or $y_k \leq pE[x_k|Y_{k-1}]$. Here the parameter p indicates a threshold value. In the first case a short position is being taken while in the later a long position is entered. In both cases, a profit can be achieved at the time the spread will revert back to its long-term equilibrium. The stochastic approach has the advantage that it reflects the mean reverting of the spread which is the main characteristic of the pairs trading techniques. What is more, using the Ornstein-Uhlenbeck process, the expected holding period as well as the expected return can be calculated explicitly using the first passage time result of the x_t process, as it is described above.

2.3.2 Distance

This is probably the most researched approach of pairs trading. Gatev et al. [24] showed that this simple method can generate profit over a big period of time. In this method, potential security pairs are sorted based on the sum of squared differences after the normalization of their prices during the formation period. As it can be clearly understood, the squared difference between the normalization of the prices series or their "distance" measures the co-movement of the pair. Then, their spread is monitored throughout the trading period and simple non-parametric threshold rules are used to trigger potential trading signals. The advantage of using a non-parametric approach helps to reduce the risk of over fitting the data. On the other hand, it is getting difficult to define a reasonable holding period or a stop-loss trigger point, due to the lack of a mathematical model. The normalised price series is its cumulative total returns index and it is calculated through the following equation:

$$P_t^i = \prod_{\tau=1}^{t} (1 + r_\tau^i), \tag{2.8}$$

where P_t^i denotes the normalised price of asset i at time t, τ is the index for all the trading days between the first trading day of the pairs formation period until day t and r_τ^i is the return of asset i.

The distance, which is the squared difference after the normalization of the price series, is calculated as follows:

$$D_{ij} = \sum_{t=1}^{N_t} (P_t^i - P_t^j)^2, \tag{2.9}$$

where D_{ij} is the squared difference between asset P_t^i and P_t^j and N_t is the total number of trading days in the pairs formation period.

The initial goal of this approach is to identify optimal pairs based on the minimum distance. Once the formation period has been passed, the trading period is getting started, where simple non-parametric rules take place. After a certain b threshold has been defined, a long position is taken, when the $P_t^i - P_t^j \geq b$, while a short position takes place at the time $P_t^i - P_t^j \leq b$. Gatev et al. (2006) proved that this strategy is profitable.

2.3.3 Co-integration

According to the co-integration framework, statistical tests are applied to identify comoving securities in a formation period. Engle and Granger [25], showed that the co-integration relationship between two stocks' prices X and Y, are identified by a stationarity test on the residual of the model below, estimated by *ordinary least squares* regression:

$$u_t = \log(Y_t) - (\alpha + \beta \log(X_t)), \tag{2.10}$$

where u_t is a stationary process with mean 0, and α and β coefficients. In the trading period, simple trading rules are being established. The key benefit of these strategies is the econometric more reliable equilibrium relationship of the identified pair. This approach will be analysed and described in more detail in the following chapters.

Chapter 3

Co-integration for Pairs Trading

Chapter 3 deals with the concept of co-integration. More precisely, basic concepts of mathematics as well as statistical concepts will be described in more detail. The chapter starts with one of the main building blocks of economical theory, describing simple terms around financial markets. Additionally, all the approaches for testing the co-integration of two time series will be introduced.

3.1 CAPM and Market Neutrality

The term market neutral tends to be one of the main features of the Pairs Trading strategy according to some authors on this topic. Market neutral means that the return of the trading strategy is uncorrelated with the market return. Hence, no matter how the market moves, the performance of the strategy is independent from it. At this point, the Capital Asset Pricing Model or CAPM is coming to us. The model links the relationship between the expected return of an individual security and the systematic risk of the market as a whole. The formula is given by:

$$E(R_i) = R_f + \beta_i(E(R_m) - R_f), \tag{3.1}$$

where the $E(R_i)$ is referred to the expected return of asset i, R_f is the risk-free-rate, β is the systematic risk or in other words, the sensitivity of the expected excess asset returns to the expected excess market returns and $E(R_m)$ is the expected return of the market.

In the above equation, the letter β plays a crucial role in our strategy. It is estimated by fitting a linear regression on asset returns against the market returns. The slope of this regression represents the parameter beta. This relationship corresponds to the so called Security Market Line or SML. The SML shows graphically the linear relationship between asset risk (beta) and its expected return. In case of a single-index model the return on all assets are related to the return on a representative index m. This index is usually taken to be a wide-ranging stock market index in the single-index model. We write the return on the ith asset as:

$$r_i = \beta r_m + \epsilon_i, \tag{3.2}$$

where r_i is the return of asset i, β is the systematic risk component, r_m is the market return and ϵ_i is the residual return. As we mentioned above, the parameter β is the slope of the regression model and it is given also by the relationship:

$$\beta = \frac{\sigma_{xy}}{{\sigma_y}^2},$$

or

$$\beta = \rho_{xy}\frac{\sigma_x}{\sigma_y}, \tag{3.3}$$

where σ_{xy} is the covariance between x and y, σ_y^2 is the variance of variable y. The equation 7.4 is coming by substituting the covariance σ_{xy} with the following relationship $\rho_{xy} = (\sigma_{xy}/\sigma_x) * \sigma_y$.

The β gives us an idea whether a stock moves in the same direction as the rest of the market, and how volatile or risky it is compared to the market. Assuming we hold a stock with $\beta > 1$, we expect that it is more volatile than the market. Hence, a stock with $\beta = 1.2$ is considered to be 20% more volatile than the market. On the one hand, if its $\beta = 0.7$, it essentially means that it is 30% less volatile than the market and hence less riskier.

Moving back to our strategy, by combining more than one assets in the portfolio, we target to get rid of the parameter β and therefore to turn our portfolio into market neutral. In this case, it doesn't matter whether the market goes up or down, the return of our portfolio stays untouchable. Therefore, in case we vanish the β, according to equation 3.2, the return is based only on the residuals term. Since the expected value of the residuals is zero, we expect a mean reverting behaviour which makes it easier for the investor to apply a 'strategy - rule' on this, creating trading signals and trying to maximize the possibilities of making profit. Such portfolios are referred to the literature as market neutral portfolios.

3.2 Stationarity

To begin with the basic structure of our analysis, it is important to mention that in the current study we cope with financial time series. Therefore, in this section we are going to analyse basic mathematical concepts on the time series analysis. Time series analysis in mathematics is a statistical technique that deals with different kinds of observations over a specific period of time. Mathematically speaking, a time series is defined as a sequential set of data points, measured typically over successive times. More precisely, we could mention that a time series is a trajectory of a discrete time stochastic process.
A time series could have different shapes. For instance, it can have an upward or downward trend or it can also look stable over time.
Mathematicians like to work with time series which have no trend. The simple reason for this, is because most statistical forecasting methods are based on the assumption that the time series can be rendered approximately stationary (i.e., "stationarized") through the use of mathematical transformations.

A stationary time series can be visualized as a form of statistical equilibrium. Stationarity is a very important concept in the field of time series analysis. Furthermore, we could mention that stationarity can be distinguished between *weak* and *strict*. However, it is rather difficult to test strict stationary and to the best of our knowledge there are only a few papers in the literature that consider testing strict stationarity. In our study, we take into account only the weak stationarity.

Definition 3.2.1 (Strict stationary) *A time series X_t is called strictly stationary if the random vectors $(X_{t_1}, ..., X_{t_n})^T$ and $(X_{t_1+\tau}, ..., X_{t_n+\tau})^T$ have the same joint distribution for all sets of indices $t_1, ..., t_n$ and for all integers τ and $n > 0$. It is written* *as:* $(X_{t_1}, ..., X_{t_n})^T = (X_{t_1+\tau}, ..., X_{t_n+\tau})^T$

Definition 3.2.2 (Weak stationary) *A time series X_t is called weakly stationary or just stationary if:*

i) $E[X_t] = \mu < \infty$, *for all t*

ii) $E[(X_t - \mu)(X_{t+h} - \mu)] = \gamma(h)$, *for all t*

Weak stationarity does not imply strict stationarity.

Every kind of analysis starts, with testing whether a time series is stationary or not. However, in most cases they fail the test of stationary. Hence some transformations methods are extensively used. Some of the most common ways to make time series stationary refer to normalization, power transform, difference transform and standardization.

3.3 Co-integration approach

Quite often the integration order of a time series is referred in parallel with stationarity.

Definition 3.3.1 (Order of Integration) *If a time series after it has been differentiated $d-1$ times $\{\bigtriangledown^{d-1} X_t\}$ is non stationary but it is stationary when $\{\bigtriangledown^{d} X_t\}$ then X_t is considered to be integrated of order d and it is denoted by $X_t \sim I(d)$.*

Co-integration is referred to a more stable and robust relationship than correlation because correlation between financial quantities in many cases seem to be unstable. Nevertheless, correlation is highly used in all multivariate financial problems. In the literature co-integration is defined in the following way.

Definition 3.3.2 (Co-integration) *The components of a vector with values $x_t = \{x_{t1}, ..., x_{tn}\}$ are cointegrated of order (d, b), denoted by $x_t \sim CI(d, b)$, if:*

i) All components of x_t are integrated of the same order d

ii) there exists a vector $b \neq 0$ such that $b' x_t$ is integrated of order $d - b$.

In a non mathematical form, co-integration is referred as an existence of a linear relationship between two time series that has constant mean and standard deviation. At all points in time, the combination between them is related to the same probability distribution. In terms of pairs trading, co-integration is linked to the process of finding a linear combination of two time series which are $I(1)$ integrated and generate a new $I(0)$ time series. The new time series, also called spread portfolio follows a mean reverting behavior. Vidyamurthy [1] provides the most cited work for this approach. He developed an univariate co-integration approach to pairs trading as a theoretical framework without empirical applications. The design of his approach looks fairly simple. The framework is based on three key steps:

- Selection of potentially cointegrated pairs, based on statistical or fundamental information.
- Testing for tradability according to a proprietary approach
- Trading rule design with non-parametric methods.

However, on one hand Vidyamurthy does not perform any co-integration test but on the other hand, the basic principle behind his framework is the idea of cointegrated pairs. We would refer that co-integration as a general concept is a useful technique for studying relationships in multivariate time series, and provides a sound methodology for modelling both long-run and short-run dynamics in a complex financial system. In order to check whether two time series are cointegrated or not, Engle and Granger (1987) proposed a simple two step procedure to test it. According to their study, a linear regression is applied on the logarithmic prices of the pair, using ordinary least squares (OLS) to estimate the coefficients and then, the residual term is tested for stationarity. The transformed asset prices by taking the logarithm of them was proposed by Vidyamurthy [1].

$$\log(Y_t) = \alpha + \beta \log(X_t) + \epsilon_t,$$

or

$$\log(Y_t) - \beta \log(X_t) = \alpha + \epsilon_t, \tag{3.4}$$

where ϵ_t is the white noise, α the intercept and β the slope or simple the hedging ratio between the assets.

The intuition behind the use of the natural logarithm of the prices, will be described below. Assuming that the prices of X and Y securities at time $t+1$ can be expressed by using the continuous rate of return r, where $r = \log(\frac{P_{t+1}}{P_t})$. Then, we get:

$$\begin{aligned} X_{t+1} &= X_t e^{r^X_{t+1}} \\ Y_{t+1} &= Y_t e^{r^Y_{t+1}} \end{aligned} \tag{3.5}$$

The spread of the values of the two securities at time t and $t+1$ are:

$$s_t = Y_t - \beta X_t, \tag{3.6}$$

$$\begin{aligned} s_{t+1} &= Y_t e^{r^Y_{t+1}} - \beta X_t e^{r^X_{t+1}} \\ &= Y_t - \beta X_t + Y_t(e^{r^Y_{t+1}} - 1) - \beta X_t(e^{r^X_{t+1}} - 1) \\ &= s_t + Y_t(e^{r^Y_{t+1}} - 1) - \beta X_t(e^{r^X_{t+1}} - 1) \end{aligned} \tag{3.7}$$

so that $s_{t+1} = s_t$ if and only if $Y_t(e^{r^Y_{t+1}} - 1) = \beta X_t(e^{r^X_{t+1}} - 1)$. If we imagine that $r^Y_{t+1} = r^X_{t+1} = r_{t+1}$, equation 3.7 turns into,

$$\begin{aligned} s_{t+1} &= s_t + Y_t(e^{r_{t+1}} - 1) - \beta X_t(e^{r_{t+1}} - 1) \\ &= s_t + (Y_t - \beta X_t)e^{r_{t+1}} - (Y_t + \beta X_t) \\ &= s_t + (Y_t - \beta X_t)(e^{r_{t+1}} - 1) \\ &= s_t e^{r_{t+1}} \end{aligned} \tag{3.8}$$

indicating that the spread value will not be constant, but widens or narrows as prices increase or decrease. Now, let's assume that the spread at time $t+1$ is defined by the natural logarithm of the prices. Hence, we have:

$$\begin{aligned} s_{t+1} &= \log(Y_t e^{r^Y_{t+1}}) - \beta \log(X_t e^{r^X_{t+1}}) \\ &= \log(Y_t) + r^Y_{t+1} - \beta \log(X_t) - \beta r^X_{t+1} \\ &= s_t + r^Y_{t+1} - \beta r^X_{t+1} \end{aligned} \tag{3.9}$$

so that $s_{t+1} = s_t$ if and only if $r^Y_{t+1} = \beta r^X_{t+1}$ and the spread will be independent of the price levels. So far, we tried to capture the idea around the fact of using price levels or the transformed prices. In fact, the co-integration approach can be applied either on the raw prices or on the transformed prices. In this thesis, the log transformation of prices is being used.

In order to test if the time series is weakly stationary, a unit root statistical test should be employed on the right hand part of the above equation. If the residual term passes the test, it allows us to construct a portfolio of assets X and Y. In the literature, there are different kind of statistical tests, which can be applied for this purpose. Each of them, has its own characteristics, its own advantages and its own disadvantages. Some of the most basic unit root tests are the Augmented Dickey-Fuller (ADF) test, the Phillips-Ouliaris test and the KPSS test. All the tests use as null hypothesis that a unit root test is present in time series sample. In the following subsection, we are going to analyse each of them.

3.4 Unit Root Test

Financial time series usually display a trend behaviour or a non stationary in the mean. As it was discussed earlier, an important task before doing any kind of analysis is to de-trend the time series. This can be achieved by different transformation methods. For instance, in *ARMA (auto regressive moving average)* modeling, the data is required to be stationary. The transformation methods of cancelling the trend depends on whether the time series is trend stationary or difference stationary. According to [21], the first case can be written as:

$$y_t = \gamma_0 + \gamma_1 t + \epsilon_t \tag{3.10}$$

where t is time and ϵ is *Gaussian* white noise. On the other hand, the second case can be written as:

$$y_t = \alpha_0 + \gamma_{t-1} + \epsilon_t \tag{3.11}$$

where ϵ is again *Gaussian* white noise. By Bhargana, nesting the previous equations, results to:

$$y_t = \gamma_0 + \gamma_1 t + u_t \tag{3.12}$$

and

$$u_t = \rho u_{t-1} + \epsilon_t \tag{3.13}$$

such that

$$y_t = \gamma_0 + \gamma_1 t + \rho[y_{t-1} - \gamma_0 - \gamma_1(t-1)] + \epsilon_t \tag{3.14}$$

where ϵ is *Gaussian* white noise. If $\rho < 1$, y_t is trend stationary process. If $\rho = 1$, y_t is difference stationary. Equation [3.14] can also be written as:

$$y_t = \beta_0 + \beta_1 t + \rho y_{t-1} + \epsilon_t \tag{3.15}$$

or

$$\Delta y_t = \beta_0 + \beta_1 + (\rho - 1)y_{t-1} + \epsilon_t \tag{3.16}$$

where ϵ is white noise, $\beta_0 = \gamma_0(1-\rho) + \gamma_1\rho$ and $\beta_1 = \gamma_1(1-\rho)$. If $\rho = 1$ then apparently $\beta_1 = 0$. Let us assume the following model:

$$\begin{aligned} y_t &= \gamma_0 + u_t \\ u_t &= \rho u_{t-1} + \epsilon_t \end{aligned} \tag{3.17}$$

then

$$\begin{aligned} y_t &= \gamma_0(1-\rho) + \rho y_{t-1} + \epsilon_t \\ y_t &= \beta_0 + \rho y_{t-1} + \epsilon_t \end{aligned} \tag{3.18}$$

with $\beta_0 = 0$ if $\rho = 1$. A *unit root test*, tests the null hypothesis $\rho = 1$ against the alternative $\rho < 1$. If there is no evidence to reject the null hypothesis, the corresponding process is considered to be non stationary. On the other hand, if the null hypothesis can be rejected given a confidence level, then the process is stationary. The following sections, which describe the mathematical point of view of different unit root methods, are based on [27].

3.5 Augmented Dickey-Fuller unit root test

The Augmented Dickey-Fuller (1984) is basically an extension of the Dickey Fuller (1979) test. The main difference between the two of them is that the Dickey Fuller test can be deployed to time series which is an Autoregressive with order 1, or in simply words it is an AR(1), while the Augmented Dickey-Fuller test can be extented to Autogressive Moving Average ARMA(p,q) processes with unknown orders. The AR(p) process is given by the following equation:

$$Y_t = \alpha + \phi Y_{t-1} + \epsilon_t, \tag{3.19}$$

where α is a constant parameter and ϵ is again a white noise. On the other hand, ARMA(p,q) process refers to the model with p Autoregressive terms and q moving-average terms and it is given by:

$$Y_t = \alpha + \epsilon_t + \sum_{i=1}^{p} \phi_i Y_{t-i} + \sum_{i=1}^{q} \theta_i \epsilon_{t-i}, \tag{3.20}$$

The Augmented Dickey-Fuller (ADF) tests the null hypothesis that the residual term after applying the OLS regression has a unit root, versus the alternative hypothesis that it is stationary. In this case, setting up a critical value usually at 5%, the null hypothesis H_0 is rejected when the p-value takes values less than 0.05 and therefore we reject the case that it is not stationary in favor of the alternative that is stationary. Briefly describing the p-value, is the probability that the statistics would be the same as or more extreme than the actual observed results under the H_0.

The corresponding test statistic is obtained by estimating an autoregression of ΔY_t on its own lags and Y_{t-1} using OLS on the following regression:

$$\Delta Y_t = \alpha + \beta t + (\phi - 1)Y_{t-1} + \sum_{i=1}^{\kappa} \gamma_i \Delta Y_{t-i} + \epsilon_t, \tag{3.21}$$

The above equation is known as the augmented Dickey-Fuller (ADF) regression. It is written in this way, such that a linear regression can be applied on Δy_t against t and y_{t-1} and test if $\phi - 1$ is different from 0. If $\phi = 1$, then we have a random walk process. If not and $\phi \in (-1, 1)$, then we have a stationary process. Under the null hypothesis, Y_t is I(1) which implies that $\phi = 1$. The ADF t-statistic and normalized bias statistic are based on the least squares estimates of 3.21 and given by:

$$\tau_{ADF} = \frac{\hat{\phi} - 1}{SE(\phi)}, \tag{3.22}$$

and

$$\tau_{ADF} = \frac{T(\hat{\phi} - 1)}{1 - \hat{\gamma_1} - ... - \hat{\gamma_k}}, \tag{3.23}$$

which is actually calculated in the same way as the usual t-statistic in an ordinary regression. The hypotheses are:

- H0 : $\phi = 1$
- H1 : $-1 < \phi < 1$

Under H_0, the distribution of the test statistic τ is, however, not t-distributed and does neither converge to a standard normal distribution. Asymptotically, τ has a so-called Dickey-Fuller distribution, which depends on the specifications of 3.21 in terms of the chosen deterministic parts, and has no closed form representation. For example, if we assume a constant but no time trend in 3.21, the limiting distribution of τ is given by:

$$\tau_{ADF} \xrightarrow{d} \frac{\frac{1}{2}(W(1))^2 - \frac{1}{2} - W(1)\int_0^1 W(x)dx}{\sqrt{\int_0^1 (W(x))^2 dx - (\int_0^1 W(x)dx)^2}}, \tag{3.24}$$

where W is a Wiener. Even though the corresponding critical values according to the above equation can easily be obtained by simulations, there is a less time consuming way provided by MacKinnon (1996), which is called response surface method.
Another important practical issue for the implementation of the ADF test is the specification of the lag length κ. If κ is too small then the remaining serial correlation in the errors will bias the test. If κ is too large then the power of the test will suffer. Ng and Perron (1995) suggest the following data dependent lag length selection procedure that results in stable size of the test and minimal power loss. First, set an upper bound $\kappa_m ax$ for κ. Next, estimate the ADF test regression with $\kappa = \kappa_m ax$. If the absolute

value of the t-statistic for testing the significance of the last lagged difference is greater than 1.6 then set $\kappa = \kappa_m ax$ and perform the unit root test. Otherwise, reduce the lag length by one and repeat the process. A useful rule of thumb for determining pmax, suggested by Schwert (1989), is:

$$\kappa = [12.(\frac{T}{100})^{1/4}], \tag{3.25}$$

where [.] denotes the integer operator returning the integer part of the corresponding argument. This choice allows $\kappa_m ax$ to grow with the sample so that the ADF test regressions 3.21 are valid if the errors follow an ARMA process with unknown order.

3.6 Kwiatkowski–Phillips–Schmidt–Shin (KPSS) test

Unlike with the Augmented Dickey-Fuller test, the null hypothesis H_0 of KPSS test assumes stationarity. Hence, similar with above, using a 5% critical value, with p-value greater that 0.05, the time series is stationary and therefore our pair is cointegrated. Additionally, in the KPSS test, the absence of a unit root is not a proof of stationarity but, by design, of trend-stationarity. This is an important distinction since it is possible for a time series to be non-stationary, have no unit root yet be trend-stationary. In both unit root and trend-stationary processes, the mean can be growing or decreasing over time; however, in the presence of a shock, trend-stationary processes are mean-reverting (i.e. transitory, the time series will converge again towards the growing mean, which was not affected by the shock) while unit-root processes have a permanent impact on the mean (i.e. no convergence over time). Mathematically speaking, KPSS test statistic is given by the model:

$$y_t = \beta t + \mu_t + \epsilon_t,$$

$$\mu_t = \mu_{t-1} + u_t, \tag{3.26}$$

where u_t is white noise with expected value 0 and standard deviation σ. The μ_t is actually a random walk. The null hypothesis as it was described before, assumes that y_t is I(0) and is formulated as $H_0 : \sigma_u^2 = 0$, which implies that μ_t is a constant. The alternative hypothesis is that the y_t process is not stationary, $H_0 : \sigma_u^2 > 0$. Although not directly apparent, this null hypothesis also implies a unit moving average root in the ARMA representation of ΔY_t. The KPSS test statistic is the Lagrange multiplier or score statistic for testing σ_u^2 against the alternative that $\sigma_u^2 > 0$ and is given by:

$$KPSS = \frac{\sum_{t=1}^{T} \hat{S}_t^2}{T^2 \hat{\lambda}^2}, \tag{3.27}$$

where

$$\hat{S}_t^2 = \sum_{j=1}^{t} \hat{\epsilon}_j, \tag{3.28}$$

where λ^2 is a autocorrelation and heteroskedasticity consistent estimator of the variance of ϵ_t. Under H0 the distribution of the test statistic depends on the inclusion of deterministic parts like a time trend. For $\beta = 0$ i.e. without deterministic time trend, and $\mu_0 = 0$, we have:

$$\tau^0_{KPSS} \xrightarrow{d} \int_0^1 W^2(x)dx, \tag{3.29}$$

where W is a Wiener process. In case we have $\mu_0 \neq 0$ and $\beta = 0$ then the distribution of the test statistic 3.28 is:

$$\tau^{(\mu_0=0)}_{KPSS} \xrightarrow{d} \int_0^1 B^2(x)dx, \tag{3.30}$$

where $B(x) = W(x) - xW(1)$ which is also called Brownian Bridge. Finally if we have $\mu_0 \neq 0$ and $\beta \neq 0$ then the distribution of the test statistic 3.28 is:

$$\tau^{(\mu_0,\beta)}_{KPSS} \xrightarrow{d} \int_0^1 B2^2(x)dx, \tag{3.31}$$

where $B2(x) = W(x) + (2\mu - 3\mu^2)W(1) + 6x(x-1)\int_0^1 W(s)ds$. The stationary test is a one-sided right-tailed test so that one rejects the null of stationarity at the $100a\%$ level if the KPSS test statistic 3.28 is greater than the $100a\%$ quantile from the appropriate asymptotic distribution 3.30 or 3.32.

3.7 Phillips–Perron test

The Phillips - Perron (1988) unit root tests differ from the ADF in the sense of the serial correlation and heteroskedasticity of the errors. To be more precise, while ADF use a parametric autoregression to approximate the ARMA structure of the errors in the test regression, the PP tests cancel out any serial correlation in the test regression. The test regression for the PP tests is as follows:

$$\Delta Y_t = \beta A_t + \gamma Y_{t-1} + \epsilon_t, \tag{3.32}$$

where ϵ_t is $I(0)$. The null-hypothesis of PP tests is that $\gamma = 0$. Compared to ADF tests, PP test is robust to general forms of heteroskedasticity in the error term ϵ_t.

Chapter 4

Theoretical implementation

Chapter 4 illustrates the framework of the development of a financial model by using two artificial stock prices, which follow a Geometric Brownian Motion and investigates the concept of statistical arbitrage on it. The intention of this chapter is also to make clear that correlation does not necessarily mean co-integration. At the beginning, statistical properties will be examined and then the performance of the strategy will be presented over a small period of time.

4.1 Geometric Brownian Motion

Geometric Brownian Motion is a building block for modelling asset prices nowadays. The model tries to capture the evolution of a stock price S_t through the following stochastic differential equation:

$$dS_t = \mu S_t dt + \sigma S_t dW_t,$$

$$\mathrm{S}_0 = s_0$$

where the term dW is a Wiener process, the parameter μ denotes the drift and the parameter σ refers to the volatility. If we rather try to modify the above equation by moving the term S_t from the right side of the equation we get the following relationship:

$$\frac{dS_t}{S_t} = \mu dt + \sigma dW_t, \tag{4.1}$$

Since,

$$\frac{dS_t}{S_t} = d(\ln S_t) = \ln(S_t) - \ln(S_{t-1}) = \ln(\frac{S_t}{S_{t-1}}), \tag{4.2}$$

therefore,

$$\ln(\frac{S_t}{S_{t-1}}) = \mu dt + \sigma dW_t. \tag{4.3}$$

The above stochastic differential equation has a unique solution which is derived using Ito's lemma. Thus the solution of the SDE is called Geometric Brownian motion with drift μ and volatility $\sigma > 0$ and takes the form:

$$S(t) = S_0 + \int_0^t \mu S(u)\mathrm{d}u + \int_0^t \sigma S(u)\mathrm{d}W(u), \tag{4.4}$$

or

$$S_t = S_0 e^{(\mu - \frac{1}{2}\sigma^2)t + \sigma W_t}, \tag{4.5}$$

where $S_0 > 0$ and $W_t \sim N(0,t)$. The above solution S_t (for any value of t) is a log-normally distributed random variable with expected value and variance given by:

$$\begin{aligned} E[S_t] &= S_0 e^{(\mu - \frac{1}{2}\sigma^2)t} E[e^{\sigma W(t)}] \\ &= S_0 e^{(\mu - \frac{1}{2}\sigma^2)t} e^{\frac{1}{2}\sigma^2 t} \\ &= S_0 e^{\mu t} \end{aligned} \tag{4.6}$$

and

$$
\begin{aligned}
Var[S_t] &= E[S^2(t)] - (E[S_t])^2 \\
&= S_0^2 e^{2\mu t - \sigma^2 t} E[e^{2\sigma W_t}] - S_0^2 e^{2\mu t} \\
&= S_0^2 e^{2\mu t - \sigma^2 t + 2\sigma^2 t} - S_0^2 e^{2\mu t} \\
&= S_0^2 e^{2\mu t}(e^{\sigma^2 t} - 1)
\end{aligned} \tag{4.7}
$$

Going back in history, this model actually corrects the Bachelier's model of 1900 (a model without the factor S_t on the right - missing the interpretation in terms of returns) which had a positive probability for negative stock prices. Since negative price on financial assets is not acceptable, Paul A. Samuelson (1973) suggested the geometric Brownian motion to model stock prices. For this, Samuelson received the Nobel Prize in Economics in 1970.

4.2 Cholesky Decomposition

In reality, we often witness asset prices not to behave in a independent way. Conversely, they tend to move up and down together. This is modelled by introducing a correlation between Brownian motions. One way to do this, is by the so-called Cholesky decomposition. Cholesky decomposition allows you to generate sample paths for two correlated Wiener processes. The basic idea is that, having a correlation matrix C for the case of two assets, which is positive-definite and symmetric, there exists a lower-triangular matrix L, such that the following relationship holds:

$$C = LL^T \tag{4.8}$$

Therefore, the correlation relationship is constructed by the following relationship:

$$\epsilon = LW, \tag{4.9}$$

where ϵ is the vector of correlated standard normal random variable, L the lower triangular matrix from the Cholesky decomposition of the correlation matrix and W independent standard normal random variable.

As a result of the above method, below we are going to construct some artificial data to check if correlation is consistent with co-integration.

4.3 Artificial data

In this section, we are going to investigate whether correlation implies co-integration by constructing artificial data.
In the first part, the two time series are assumed to follow a geometric Brownian motion. The construction is given as follows. We simulate the behaviour of one asset, assuming it follows a geometric Brownian motion. Then, the second asset is built by shifting the first asset above by some positive value and adding also white noise. The random noise which has been introduced, is assumed to follow a normal distribution. The parameters of the geometric Brownian motion model are $\mu = 0.04$, $\sigma = 0.2$ and initial stock price equals to $S_0 = 100$. In the second part, the construction of the two correlated time series will be done by using the Cholesky decomposition, as it was described in the previous section. Below, the correlation matrix is defined as $\Sigma = \begin{bmatrix} 1 & 0.9 \\ 0.9 & 1 \end{bmatrix}$, drift and volatility are $\mu_1 = 0.01$, $\mu_2 = 0.03$, $\sigma_1 = 0.05$, $\sigma_2 = 0.03$ and the initial price for both of them is $S_1 = S_2 = 100$. The time framework in the below graphs is considered to be 6 months, starting in April and ending up in October. Each graph is divided into two parts. The above part presents the prices of time series and the below part the spread of them.

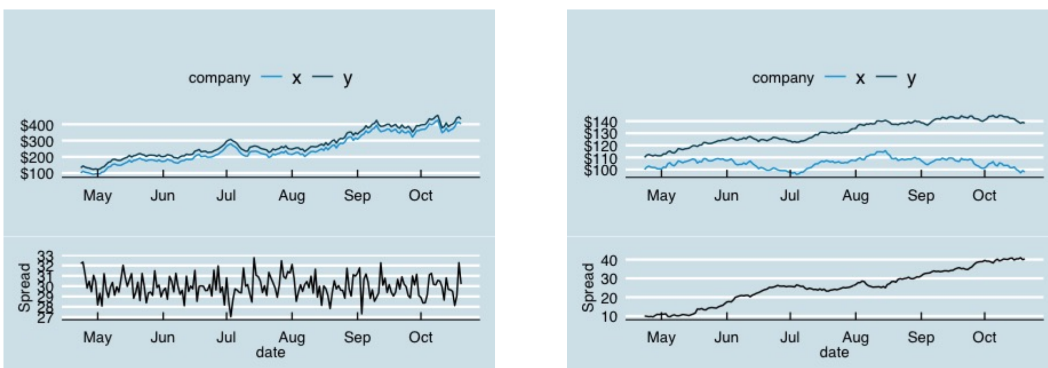

Figure 4.1: Artificial data with a degree of correlation

As it is shown in figure 4.1, in the left hand side plot, the prices of both companies move together across the corresponding period. Furthermore, in the lower part of this figure, we can observe the spread of the prices, which looks quite stable around a constant mean. In the right hand side plot, the prices seem to be correlated, since there are periods that the prices move simultaneously up and down. However, the spread of the prices displays an upward trend and it doesn't look stationary around a mean, as it can be seen in the right part of figure 4.1.

At this point, we can proceed with a statistical analysis, in order to detect and to make clear whether correlation implies co-integration. Below, the tables show correlation results, as well as, the p-value of augmented Dickey–Fuller test. The left column represents the data set which is captured on the left hand side of the figure 4.1, while data set 2 refers to the right hand side.

	Data set 1	**Data set 2**
Correlation of prices	0.999	0.877
Correlation log-returns	0.985	0.923
ADF test (p-value)	<0.01	0.84

According to the results, a strong relationship is shown between variables x and y. Their correlation takes a value of almost 1. Meanwhile, the augmented Dickey–Fuller test shows that we have strong evidence to reject the null hypothesis of non stationarity. On the other side, data set 2 shows also a strong correlation between the variables. However, the p-value of the augmented Dickey–Fuller test seems to be high and therefore it is hard to confirm a stationarity of the spread of the variables. To sum up, we showed that although two random assets could be highly correlated, that does not mean that they are also co-integrated. In the following section, we are going to examine the Engle-Granger two step procedure, applied on the data set 1 and investigate how this could work in terms of pairs trading.

4.4 Engle-Granger procedure

As it was described previously, the spread of data set 1 does not follow a unit root process. Therefore, we have strong evidence that the stock prices co-move and we are going to apply the two time step procedure, which was introduced by Engle-Granger (1987) and described in the previous chapter. At a quick glance, we can observe that both stochastic processes are non stationary, showing both an upward trend. Thus we can run a least square (LS) regression on the observed data. Thus, we can estimate the unknown parameters of the linear regression model. The model is as follows:

$$y_t = \hat{a} + \hat{\beta}x_t + \hat{\epsilon}_t \tag{4.10}$$

Solving by $\hat{\epsilon}_t$ the relationship 4.10 takes the form:

$$\hat{\epsilon}_t = y_t - (\hat{a} + \hat{\beta}x_t) \tag{4.11}$$

The above equation shows essentially the evolution of the residuals, which are the estimated values of error. In case that y_t and x_t series are co-integrated, that means that the residuals are I(0), or mathematically speaking stationary. To confirm this case, we have to apply a Dickey Fuller test on the residuals. According to section 3.4, the main idea is to apply a linear regression on the change of the residual term based on its lagged value. Then, a t-statistic test is employed in the $\phi - 1$ term and it is compared with the Dickey Fuller distribution. If t is less than a particular value of the Dickey Fuller distribution, we could reject the null hypothesis that the residual term is I(0). Based on the data set 1, the p-value is less than the 5% and thus we reject the null hypothesis. Hence, our pair is co-integrated.

4.5 Trading rule

In the previous section, we concluded that the data set 1 passes the co-integration test. Now, it is time to see how we could profit from the linear relationship that they have. According to the regression equation, the coefficient β, or in other words the hedging ratio, takes a value which is close to 1 (0.998). Hence, the theoretical portfolio could be constructed through the following equation:

$$\text{Portfolio} = y_t - \beta x_t$$

or

$$\text{Portfolio} = y_t - x_t \tag{4.12}$$

Its mean reverting behavior can be seen in figure 4.2. What is more, the red line indicates the average value, while the up and down dashed lines show, the ± 1 standard deviation from the mean.

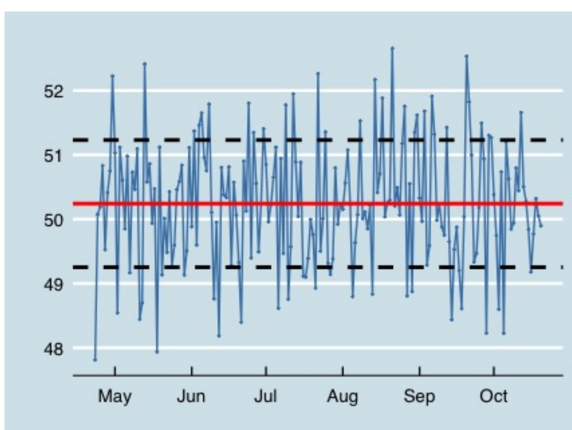

Figure 4.2: Spread of the theoretical pair

Since the spread in equation 4.2 has a mean reverting behavior, we can easily apply some rules or in other words, a trading strategy. In pairs trading framework, the rule is considered as entering a long position when the spread exceeds the lower standard deviation and a short position when it crosses the upper standard deviation. All positions close when the spread reverts to a certain threshold. Usually this threshold refers to the mean. However, the closing threshold, as well as the two thresholds that a position is being opened, is part of an optimization method where the thresholds are chosen in a way such that the overall return is maximized. In figure 4.3, we have created some trading signals. Sometimes the fact that we refer to the spread and not on the prices of our pair, can be confusing . In this example, a long position on the spread means, we buy one unit of y_t and sell one unit ($\beta = 1$) of x_t. Similarly, entering a short position on the spread is translated in selling one unit of y_t and buying one unit ($\beta = 1$) of x_t. At this point, it is crucial to mention that we open a position because we have only statistical evidence that in a future time the spread will revert to a long equilibrium. Therefore, it is wise to set up also some additional rules, which minimise the risk that the spread for some reason will never return to its mean.

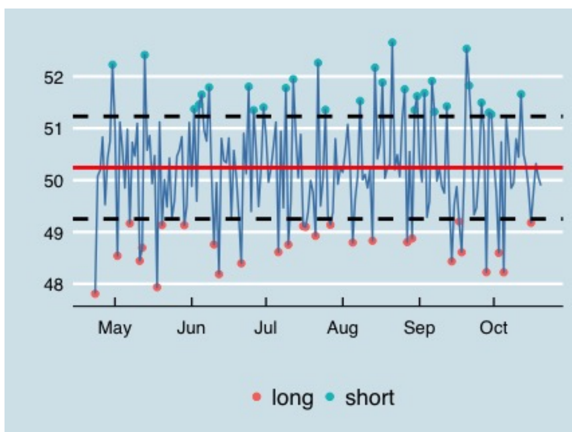

Figure 4.3: Trading signals on the portfolio

Applying the rule which has already been described above, we could extract some useful information. The main scope for every strategy is to test its performance. One of the most popular statistics is the cumulative return of a portfolio over a period of time. Speaking mathematically, cumulative return is the aggregate return that the portfolio has gained or lost over a period of time. It is defined as:

$$\text{Cum.Return} = \frac{\text{Current.price} - \text{Initial.price}}{\text{Initial.price}} \tag{4.13}$$

In figure 4.4, we can see the cumulative return of the pairs trading strategy, compared with the cumulative return of the time series x and y. It can be seen that the portfolio

, outperforms the two time series. In the following chapter, we are going to test this strategy on real financial time series.

Figure 4.4: Profit and loss based on the cumulative returns

Chapter 5

Kalman filter in Statistical Arbitrage

Chapter 5 deals with the relationship between the Kalman filter and the Statistical Arbitrage framework. The Kalman filter is going to be analysed in more detail. Focus has been mainly on the theoretical part, where the derivation will also take place.

5.1 Review of Kalman Filter

Kalman filter, according to the literature is an efficient recursive filter, which estimates states of a linear dynamic system from a time series of observations. Combined with the linear quadratic regulator, the Kalman filter solves the linear-quadratic-Gaussian problem, that is one of the most fundamental optimal control problems according to Ali Hirsa [14].

As stated in Wikipedia, the name comes from the Hungarian Rudolf E. Kálmán, although, Thorvald Nicolai Thiele and Peter Swerling developed a similar algorithm earlier. The first implementation of this method was at that time, when Kálmán visited the NASA Ames Research Center. Then Schmidt saw the applicability of Kálmán's ideas to the nonlinear problem of trajectory estimation for the Apollo program leading to its incorporation in the Apollo navigation computer.

Nowadays, the Kalman filter is extensively used in various fields. As reported by Wikipedia, we can observe many applications in technology, like NASA's trajectory estimations, the US Navy's ballistic missile submarines, radar technology, and more recently—global positioning systems (GPS), satellite tracking and robotics. What is more, the Kalman filter is a widely applied concept in time series analysis which is used more specifically in fields, such as signal processing and econometric. One of the biggest advantages of this method is that it can cope well with multi-dimensions in both the state and observation matrices. What is more, the Kalman filter uses a set of equations to iteratively measure successive observations with increasing accuracy, using only the previous estimate and the current estimate. Therefore, it can be described as a very efficient algorithm because it doesn't require any memory since we dynamically update our prediction for each step using only the current information. So, the method combines two extremely useful properties which are working fast in high dimensions.

Now, the question is how this dynamic system is related to statistical arbitrage. In real life, as we have investigated into previous sections, it is difficult to find financial time series, which have stable over time co-integrated properties. Instead, the slope or in other words the hedging ratio of the pair is changing across the time. So, one way to cope with both uncertainty and dynamism in our decisions is to use the Kalman filter for parameter estimation. The Kalman filter is a state space model for estimating a "hidden" state process by using observations of related variables and models of those relationships. Hence, the Kalman Filter consists of a state space with a fully observed process and one unobserved or hidden state. A process that the probability density function of the process is available in an integrated form is called a fully observed process. An observed process is a process that the probability density function of the process is available in an integrated form. The estimation procedure for these processes is done via maximum likelihood estimation, which we will describe below. On the other hand, except from observed processes, there are some other, called partially observed, in which the density function is not available in an integrated form. To get a density in an integrated form, the partial observed process is being conditioned on a set of parameters. This set of parameters form the hidden state process. In terms of partially observed processes, when we deal with a time series, the goal is to calculate for each day the hidden process such that the hidden state will fit with the true observation as better as it can. The procedure is called filtering. Another issue on this problem is how to choose the "best" parameters on the set of our model so that the hidden process will match with the observed data. The process is called parameter estimation and it is done by using a set of parameters that maximizes the likelihood function or yields the smallest mean square root error.

5.2 A state space model

A state space model refers to a model where the parameters follow a random walk. As stated in [21], we can start describing the Kalman filter concept by taking a general model, from which many other special cases' model can be obtained. Hence the general model is as follows:

$$\begin{aligned} &\text{Measurement equation:} && y_t = \beta_t x_t + u_t && (5.1)\\ &\text{Transition equation:} && \beta_t = \beta_{t-1} + v_t, \quad t = 1, ..., n && (5.2) \end{aligned}$$

where,

$$\begin{pmatrix} u_t \\ v_t \end{pmatrix} \sim \mathcal{N}\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_u^2 & 0 \\ 0 & \sigma_v^2 \end{pmatrix}\right]$$

In 5.1 and 5.2, the first equation refers to the measurement equation while the latter refers to the transition equation. What is more, y_t is a $n \times 1$ vector, x_t is a $n \times m$ matrix and β_t is a $m \times 1$ vector of coefficients. Also, in the majority of applications of these models, the variances σ_u^2 and σ_v^2 are assumed to be known. By substituting the 5.2 equation into the 5.1 we get:

$$\begin{aligned} y_t &= (\beta_{t-1} + v_t)x_t + u_t \\ &= \hat{\beta}_{t-1}x_t + (\beta_{t-1} - \hat{\beta}_{t-1})x_t + x_t v_t + vt \\ &= \hat{\beta}_{t-1}x_t + w_t \end{aligned} \tag{5.3}$$

where $\hat{\beta}_t$ is the estimator of β_t based on observations up to t and w_t is the one-step ahead prediction error. If $var(\hat{\beta}_{t-1}) = \lambda_{t-1}$ then $var(w_t) = f_t$ where $f_t = x_t^2\lambda_{t-1} + x_t^2\sigma_v^2 + \sigma_u^2$. Then maximizing the likelihood is equivalent to maximizing :

$$-\frac{1}{2}\sum\left[\log f_t + \frac{w_t^2}{f_t}\right]$$

5.2.1 Process of Kalman Filter

According to Ali Hirsa [14], the Kalman filtering process can best be described as a three step process of prediction, observation, and correction. In the prediction step, we predict the next system state based on our knowledge of the current system state. Along with it, we also estimate the error in our prediction. This completes the prediction step. Then the new data arrives and the system would now have transitioned to the new state. Similar to the prediction step, we also estimate the error associated with the true observation. The observation along with an estimate of the error constitutes the observation step. We now have two estimates for the states involved: one based on our prediction and the other based on our observation. The final step consists of the correction of our prediction based on the observation which has already arrived. This is therefore called the correction step. This reconciled estimate of the system state from the correction step is the final estimate of the current system state. After this step, the process is repeated at the next time step making the Kalman filter a recursive prediction correction method.

Mathematically speaking as stated by Pole [2], the expected value of β at time t given all the information up and till time $t-1$ by $\hat{\beta}_{t|t-1}$ and the expected value of β at time t given all the information up and till time t by $\hat{\beta}_{t|t}$. In a similar way the expected value of y at time t given all the information up and till time t is defined by $\hat{y}_{t|t-1}$. Given the initial values $\hat{\beta}_{t-1|t-1}$ and $R_{t-1|t-1}$, which is the covariance between the observed β_t and the predicted $\hat{\beta}_{t|t-1}$, at time $t-1$, we can start with the first prediction step. The whole process is described by the following equations:

$$\hat{\beta}_{t|t-1} = \hat{\beta}_{t-1|t-1} \qquad \text{state prediction} \tag{5.4}$$

$$R_{t|t-1} = R_{t-1|t-1} + V_w \qquad \text{state covariance prediction} \tag{5.5}$$

$$\hat{y}(t) = \hat{\beta}_{t|t-1} x(t) \qquad \text{measurement prediction} \tag{5.6}$$

$$Q(t) = x'(t) R_{t-1|t-1} x(t) + V_e \qquad \text{measurement variance prediction} \tag{5.7}$$

where R is a 2×2 matrix as was explained above and Q is the variance of the forecast error $y(t) - \hat{\beta}_{t|t-1}x(t)$. Now, the next step is to update the forecast, since the new data has arrived.

$$\hat{\beta}_{t|t} = \hat{\beta}_{t|t-1} + K(t) \times e(t) \qquad \text{state update} \tag{5.8}$$

$$e(t) = y(t) - \hat{\beta}_{t|t-1} x(t) \qquad \text{prediction error} \tag{5.9}$$

$$R_{t|t} = R_{t|t-1} - K(t) \times x(t) \times R_{t|t-1} \qquad \text{state covariance update} \tag{5.10}$$

where K is called the Kalman gain and it is the correction term applied to the one-step ahead prediction error $e(t)$. The Kalman gain is given by:

$$K(t) = R_{t|t-1} \times x(t) \times \frac{1}{Q(t)}$$

5.2.2 Derivation of Kalman filter

As stated in [21], to derive the Kalman filter given by equation of state update and covariance update, we shall use the following fact, based on normal regression theory. Suppose:

$$\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} \sim \mathcal{N}\left[\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}\right]$$

Then the conditional distribution of y_1 given y_2 is normal with mean $\mu_1 - \Sigma_{11}\Sigma_{22}^{-1}(y_2 - \mu_2)$ and variance $\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$.

We shall use this fact to derive the distribution of $\hat{\beta}_{t|t}$ conditional on y_t. First note that:

$$\begin{aligned} \hat{\beta}_{t|t} &= \hat{\beta}_{t|t-1} + \hat{\beta}_{t|t} - \hat{\beta}_{t|t-1} \\ y(t) &= \hat{\beta}_{t|t-1}x(t) + x(t)(\hat{y}(t) - \hat{\beta}_{t|t-1}) + e(t) \end{aligned}$$

Hence, we have a multivariate normal distribution as follow:

$$\begin{pmatrix} \hat{\beta}_{t|t} \\ y_t \end{pmatrix} \sim \mathcal{N}\left[\begin{pmatrix} \hat{\beta}_{t|t-1} \\ \hat{\beta}_{t|t-1}x_t \end{pmatrix}, \begin{pmatrix} R_{t|t-1} & R_{t|t-1}x_t \\ x_t R_{t|t-1} & Q_t \end{pmatrix}\right]$$

Taking an estimator of $\hat{\beta}_{t|t}$ as the conditional mean $E[\hat{\beta}_{t|t}|\hat{y}_t]$ and using the above result we get the Kalman filter equations of state update and state covariance update. The above analysis is based on the assumption that $E[\hat{\beta}_{t|t}|\hat{y}_t]$ is the best estimator of β given the observations up to and including time t. Also, the Kalman filter estimator of β depends on the normality assumption. If the disturbances in the state-space model are not normally distributed, it is no longer true, in general, that the Kalman filter yields the conditional mean of the state vector. However, even if the disturbances are not normal the estimator $\hat{\beta}_{t|t}$ given by the Kalman filter is the minimum mean-squared-error linear estimator of β based on the observations up to and including time t.

5.3 Results of Kalman filter

At this point of our analysis, the fundamental question arises on how well the Kalman filter model captures the hedging ratio. Applying all the previous steps and having established the initial values of $\hat{\beta}_{1|0} = R_{0|0} = 0$ we can present our results. In terms of variances V_w and V_e, Rajamani and Rawlings (2007, 2009) introduced autocovariance least squares method in order to estimate these covariances. According to the literature, V_w refers to a 2×2 matrix taking values $\frac{\delta}{1-\delta}I$, with I to be the 2×2 identical matrix. The parameter δ takes values between 0 and 1. If it takes the value 0, then our model turns into an OLS regression model with fixed β. On the other hand, if $\delta = 1$, the parameter β is going to volatile a lot from time step to time step. Our estimate follows that $\delta = 0.0001$ and $V_e = 0.001$. In the following graph, we can clearly observe how the hedging ratio fluctuates across this period of time. It is obvious that the slope is changing and it is not remaining static like in the co-integration method.

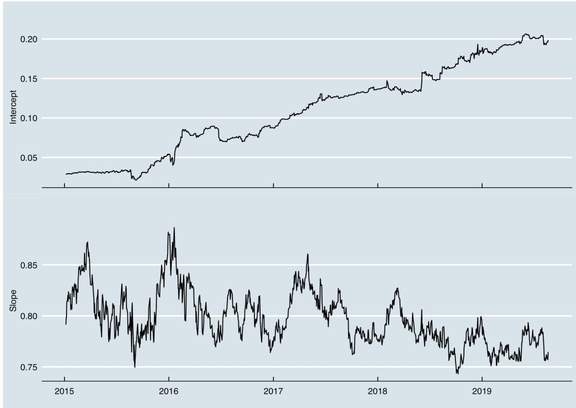

Figure 5.1: The intercept and the slope obtained by Kalman filter

In terms of the prediction against the actual data, the Kalman filter generates good results, as it can be seen in the next plot.

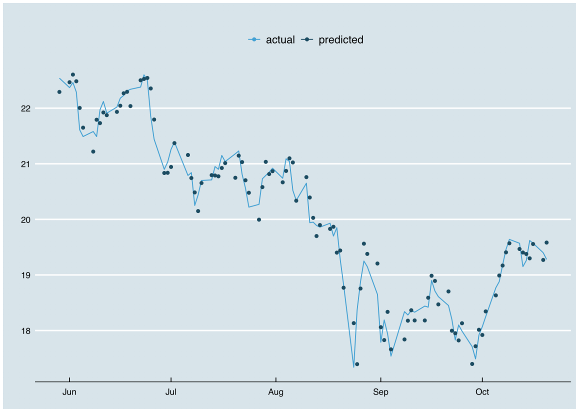

Figure 5.2: Prediction performance of Kalman filter

Furthermore, in the next chapter, we will examine the profitability of the Kalman filter along with the classical approach.

Chapter 6

Empirical Analysis

Chapter 6 focuses on the application side of statistical arbitrage method. It starts with a short explanation of the data that will be used. Then, co-integration is tested on the training data. Focus is placed on the avoidance of having a bias estimator (hedging ratio). The goal of this chapter it to figure out under different scenarios, the performance of this strategy. Furthermore, a performance test based on the dynamic estimate of the hedge ratio in a pairs trading strategy is deployed.

6.1 Data

It is very rare nowadays to find a pair of assets or even a group of assets on the market, which fulfill all the statistical properties, so, at the end we would have some evidence that they don't drift apart to each other over a period of time. For the purpose of this study a pair of exchange traded funds-ETFs is being used. But what actually means ETF?
An exchange traded fund (ETF) is a type of security that involves a collection of securities, such as stocks that often track an underlying index, although they can invest in any number of industry sectors or use various strategies. ETF's are in many ways similar to mutual funds; however, they are listed on exchanges and ETF shares trade throughout the day just like ordinary stock. Exchange traded funds is a good reason for someone who is willing to expand the diversification of his/her portfolio. Therefore, the two ETFs are the iShares MSCI Australia ETF and the iShares MSCI Canada ETF. Below, a small description of them according to Bloomberg is done.

IShares MSCI Australia ETF is an exchange-traded fund incorporated in the USA. The ETF tracks the performance of the MSCI Australia Index. The ETF holds predominantly large-cap stocks from Australian companies. The ETF weights the holdings using a market capitalization methodology and re-balances quarterly.

IShares MSCI Canada ETF is an exchange-traded fund incorporated in the USA. The ETF tracks the performance of the MSCI Canada Custom Capped Index. The ETF holds large and mid-cap stocks. Its investments are focused in Canada. The ETF weights the holdings using a market capitalization methodology and re-balances quarterly.

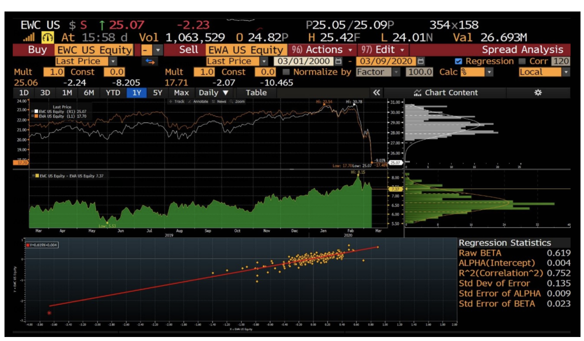

Figure 6.1: Raw data of EWA US Equity and EWC US Equity from Bloomberg

6.2 Analysis

As we have done in chapter 3 in the case of artificial data, we will follow the same steps in order to find out initially whether our real pair fulfill the statistical requirement of co-integration. For this task, the parameter β or the hedging ratio in other words, should be extracted by applying a simple linear regression model on the pair. The parameter indicates the slope of the regression line. However the way we split the data into train and test data, seems to be very crucial. To avoid having a bias estimator at the end, there are different ways of validation in the machine learning field, which will be described in the following section. Our data set, consists of daily closing prices covering a range of almost 20 years (2000-2019) period.

6.2.1 Cross-Validation

Data scientists want to fit training models to data that will do a good job of predicting future, out of sample data points. In the meantime it is getting extremely important to avoid estimating biased estimators that create at the end an over fitting model. As in the over fitting case, under fitting is also another issue, which refers of not capturing enough patterns in the data. As a result, the performance not only of the training set but also of the test set is quite poor. For all these reasons, there are several techniques out there, which try to prevent any kind of risk model.

According to the most typical case, the data is split into training and test data set. Although, the proportion of the splitting data varies a lot, a 70-30% or 80-20% is quite common and reasonable. Additionally, having a lot of data is quite useful for this method.

Another very well known approach is the so called K-fold. The main idea, since we do not hold endless data is to tackle the problem of under fitting and over fitting as well. By reducing the training data, we risk losing important patterns in the data set, which in return increases the error induced by bias. So, the method suggests to split the data set into k folds. Then going iteratively k times, every data set is used k times for validation purpose and $k-1$ times for training purposes. The number of k accounts most of the times either as 5 or as 10. This procedure reduces significantly the under fitting since we are using most of the data for fitting, and also reduces significantly the over fitting as most of the data is also being used in the validation set.

6.2.2 Method

After consideration, in this study, a combination of the above methods is going to be used as one method in this study. The method is known as walk-forward analysis. To describe the walk-forward method in a better way, each time approximately 15 years of data is going to be used for training and 3 months data for testing. Into the training set, additionally a 1 year rolling window is also applied, taking each time the average of the β parameter. On the next step, the data goes 3 months ahead and the procedure is repeated. The whole process is only done, when we run out of testing data. At the end, the test data are being aggregated and the whole performance can be examined. Trying to visualize our thought regarding the cross validation, the result can be seen in figure 6.2.

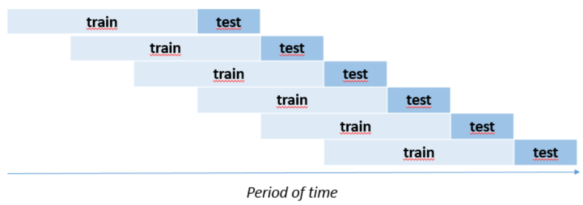

Figure 6.2: Cross-validation, walk-forward analysis

Furthermore, apart from the walk-forward process a version of a simple splitting method (75-25%) will be tested and compared with the others. In this approach, a one year rolling regression will be applied on the train data set. Then, the coefficient β (hedge ratio) is extracted by calculating the average of the coefficients β and finally the spread is built. Finally, as in case of simple splitting data, a monthly rolling regression will be applied on the test set in order to capture better the behavior of the coefficient β (hedge ratio). Additionally, we compare the above methods, which are coming from the co-integration concept, with something more dynamic, called Kalman filters. As we have already analysed in the previous chapter, Kalman filter is an alternative method of parameter estimation. Since, we have a quite big amount of data, the performance of statistical arbitrage will be examined during the last 4.5 years, counting from 2015 till today.

6.3 Back-test

In case we do not want to wait for the markets, in order to test our strategies, this is where backtesting comes in. Backtesting is the concept of testing your ideas on historical data. Historical performance certainly doesn't guarantee future returns, but it can go a long way towards growing confidence in your strategy. With backtesting, you simply apply your strategy at a certain date, and test how that strategy would perform into the future. Variance always plays a role, and we must be careful to not grow a biased view during our backtesting. It is quite wise to test your model on as much data as it is possible. Fortunately, there are petabytes of data on transactions, customers, bills, money transfers, prices and so on. Hence, there is a plethora ways to feed a model. Then, if the results look promising, it is time to evaluate how the strategy matches up on live returns by simulating trades in real time, using an automated system.

6.3.1 Statistical results

This section focuses on the statistical tests of our chosen pair. Even though, EWA and EWC are funds tracking different kind of MSCI indices, we expect not only a high degree of correlation but also a possible co-integration. To have a better view of our pair, we plot below the development of closed prices over a period of 20 years. The left part shows the closed prices as they are, while the right part shows the prices after normalization with z-scores.

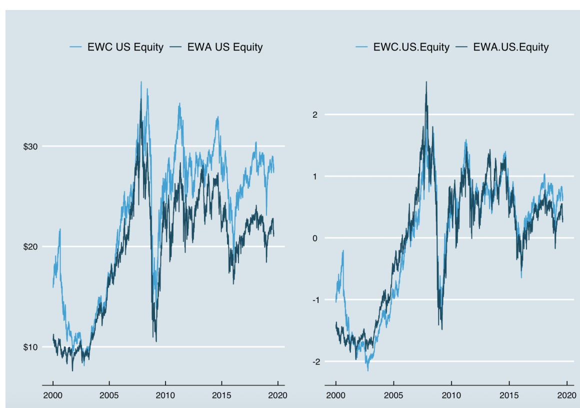

Figure 6.3: Price development

We can observe in figure 6.3 a strong relationship between the prices and this is proved by the high degree of correlation which takes the value of 0.951. The correlation of logarithmic returns accounts to 0.72. The following graph proves the corresponding numbers.

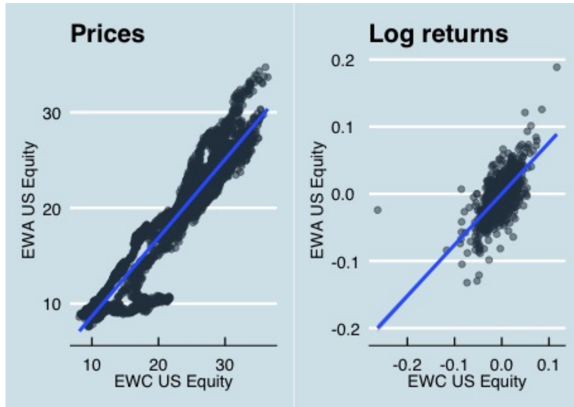

Figure 6.4: Correlation of prices (left) and logarithmic returns (right)

What is more, to have a better overview of the returns, the monthly return in figure 6.5 for the entire period for each asset was calculated. Below, we can see the output, which presents big spikes during the financial crisis in 2008. Also, a slightly bigger volatile period around 2011 can be witnessed. This could be an effect of the debt crisis in 2009-2011 in Europe, which hit also the markets across the world.

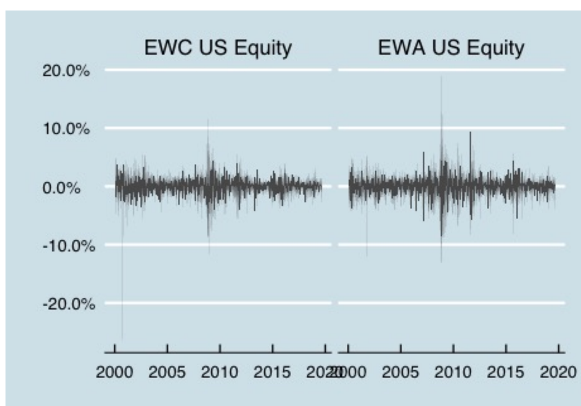

Figure 6.5: Monthly returns

6.3.2 co-integration analysis

In this step, it is crucial to test if our pair passes a co-integration test or not. Economically speaking, there is strong evidence that there is a relationship between them, as they are both tracking the performance of MSCI indices. Firstly, both of them are ETFs. Second of all, iShares is a family of exchange-traded funds (ETFs) managed by BlackRock, which is an American global investment management corporation based in New York City. Both EWA and EWC try to track indices from developed countries like Canada and Australia. For the above reasons, we proceed with a co-integration test. For that, the two step Engle and Granger (1987) method will be used. Additionally, Kwiatkowski-Phillips-Schmidt-Shin (KPSS) and the Phillips-Perron (PP) tests are used in order to check whether all tests confirm a stationarity or not. The results are on the following table.

Uni-root test	**Statistic**	**P-value**
Augmented Dickey Fuller (ADF)	-3.384	0.04374
Phillips-Perron (PP)	-22.894	0.03325
Kwiatkowski-Phillips-Schmidt-Shin (KPSS)	2.8762	0.01

Since, the p-value in all cases takes a value less than 0.05, having chosen a 95% confidence level, we conclude to reject the null hypothesis of non stationarity in favor of the alternative hypothesis. This essentially means that the confidence interval covers the true value in 95 of 100 studies performed. Right now, we have enough confidence and also strong statistical evidence about a potential portfolio with a mean reverting behaviour. The two step Engle and Granger method with an average of one year rolling window proposes a β of 0.9176 . Therefore, the portfolio follows the equation:

$$\text{Portfolio} = EWC - 0.9176 \cdot EWA \tag{6.1}$$

The spread over the testing period (2015-2019) using the equation 6.1 is as follows:

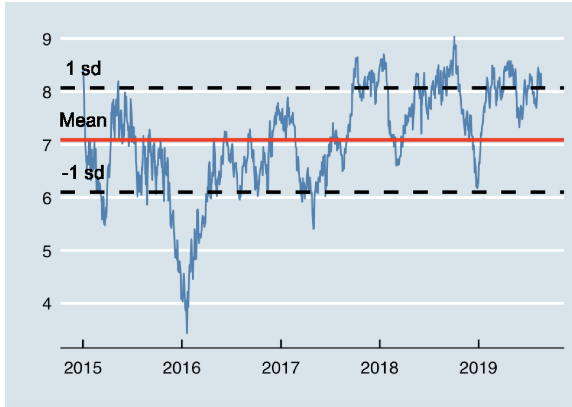

Figure 6.6: Spread of the portfolio during the period 2015 - 2019. Horizontal lines refer to the average of the spread along with ± 1 standard deviations

6.3.3 Performance analysis

In the previous section, the portfolio of pairs has already been defined. The parameter β has a positive value and therefore we will simple define the trading rule.
For each case of the cross-validation procedure, we define an upper and a down bound across the spread portfolio. The spread portfolio, as it was discussed above, has the form, "Portfolio $=$ EWC $- \beta \cdot$ EWA". According to the most well known rule of the statistical arbitrage strategy, a short position is taken on the overvalued asset, while a long position is taken on the undervalued asset. But how is the undervalued or overvalued asset identified? This can be seen by the behaviour of the spread. When the spread simple exceeds the upper bound, the investor should take a short position on the EWC US Equity and meanwhile a long position on β units of EWA US Equity. The same rule is applied on the other way round, which means, when the spread goes below the down bar the investor goes long the EWC US Equity and short β units of EWA US Equity. This will bring a profit at the time the spread returns to its mean and the positions are being closed. During our back-test, we are going to test the performance of this strategy by using different thresholds. Hence, ± 1 standard deviation, ± 1.5 standard deviation, ± 2 standard deviation will be investigated. Along with the performance measures, we present also some risk metrics. These are the Sharpe ratio and the maximum drawdowns. To define a drawdown, it describes how much an investment or trading account is down from the peak before it recovers back to the peak. Drawdowns are important for measuring the historical risk of different investments, comparing fund performance, or monitoring personal trading performance. On the other hand, the Sharpe ratio is a very well known measure and it refers to the average return, earned in excess of the risk-free rate per unit of

volatility or total risk. In simple words, it helps investors to understand the return of an investment compared to its risk. The bigger the Sharpe ratio is, the better performance we have. All the results, are located on the following tables. For each table, an aggregate figure of the performance is presented.

Comparing results of Pairs Trading methods with Index and individual securities. The upper and down threshold of pairs trading strategy has been defined to 1 standard deviations from the mean. The period of back-test starts in Jan 2015 and ends in Oct 2019.

Pairs Trading	Method	Annual return	Annual sd	Max Drawndown	Sharpe
1	Simple split	6.49%	13.80%	0.2169	0.4700
2	Rolling window	8.17%	10.76%	0.2748	0.759
3	Walk-forward	16.03%	12.52%	0.1426	1.2797
4	Kalman filter	6.90%	10.37%	0.1588	0.665
	S&P 500	6.7%	13.67%	0.2025	0.4899
	EWA	-2.49%	18.99%	0.3522	-0.1314
	EWC	-1.98%	15.74%	0.3781	-0.1257

In the above table it can be clearly seen that the **walk-forward** procedure outperforms all the other methods. Its Sharpe ratio looks quite promising as well. Furthermore, we can see that the **simple split** method is very close to the benchmark index S&P 500 but it cannot outperform it. Below, a clear picture of all the methods during the testing period is being plotted.

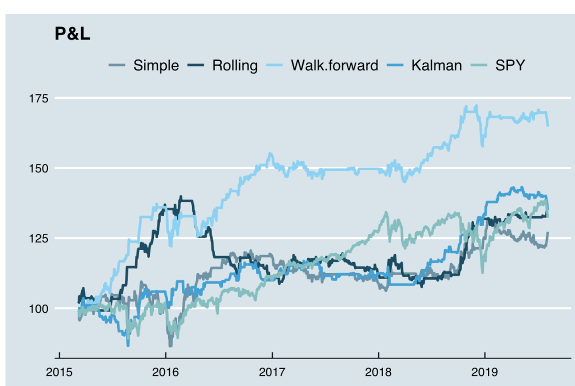

Figure 6.7: Cumulative returns of the methods using as a threshold ±1 standard deviation

Comparing results of Pairs Trading methods with Index and individual securities. The upper and down threshold of pairs trading strategy has been defined to 1.5 standard deviations from the mean. The period of back-test starts in Jan 2015 and ends in Oct 2019.

Pairs Trading	Method	Annual return	Annual sd	Max Drawndown	Sharpe
1	Simple split	4.27%	12.08%	0.2017	0.3535
2	Rolling window	5.11%	7.86%	0.224	0.6502
3	Walk-forward	21.43%	10.64%	0.0952	2.0139
4	Kalman	5.87%	7.07 %	0.0887	0.8295
	S&P 500	6.7%	13.67%	0.2025	0.4899
	EWA	-2.49%	18.99%	0.3522	-0.1314
	EWC	-1.98%	15.74%	0.3781	-0.1257

Similarly to the previous table, the **walk-forward** procedure outperforms all the others. It seems actually that in this case, it generates a magnificent annual return of 21.5%, which is followed by a large Sharpe ratio as well. However, the other methods generate poor returns, which makes sense since we have increased the distance from its mean and there are less trading signals. What is more, the **simple split** method seems to trigger very few signals and therefore its performance does not seem attractive. The following figure verifies the above results.

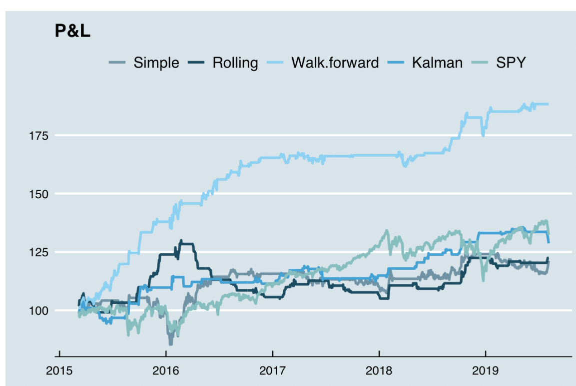

Figure 6.8: Cumulative returns of the methods using as a threshold ±1.5 standard deviation

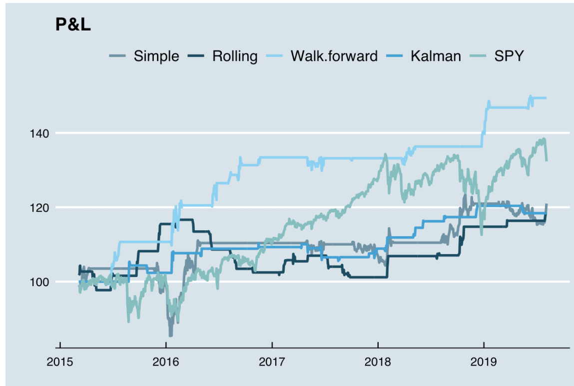

Figure 6.9: Cumulative returns of the methods using as a threshold ±2 standard deviation

Comparing results of Pairs Trading methods with Index and individual securities. The upper and down threshold of pairs trading strategy has been defined to 2 standard deviations from the mean. The period of back-test starts in Jan 2015 and ends in Oct 2019.

Pairs Trading	Method	Annual return	Annual sd	Max Drawndown	Sharpe
1	Simple split	3.82%	10.31%	0.1884	0.3701
2	Rolling window	4.07%	6.17%	0.1612	0.6592
3	Walk-forward	11.80%	7.74%	0.0629	1.5247
4	Kalman filter	3.85 %	4.61%	0.04673	0.8347
	S&P 500	6.7%	13.67%	0.2025	0.4899
	EWA	-2.49%	18.99%	0.3522	-0.1314
	EWC	-1.98%	15.74%	0.3781	-0.1257

As in the previous tables, the **walk-forward** method was the best performing, while the **simple split** method was the worst. Nevertheless, it is clear that by trading some combination of the pair, it is definitely better than trading the individual securities, which both present a negative return. The [6.9] plot confirms the above table.

Chapter 7

Pairs trading in a volatile market

Chapter 7 describes different risk measures for a portfolio of co-integrated pairs during the COVID-19 period. Furthermore, it answers the question whether pairs trading is a neutral strategy. In this chapter, real financial data, obtained from YAHOO is used in order to proceed with the analysis.

7.1 Recession 2020

2020 is a year that has been characterised as the most challenging year after the financial crisis in 2008. At this time, it didn't start as a financial crisis. It was a virus, started from China and spread to the West within a couple of months. At a high speed, the virus affected not only the health system of the countries but also damaged seriously the financial sector. In order to tackle the problem, governments announced trillions in USD of stimulus packages in order to boost the economies. As a result, the global debt has never been higher, bringing a lot of uncertainty. Meanwhile, the macroeconomic data is looking terrible. As it can be easily understood the market is described by a high volatility.

Under these circumstances, we believe that after we have tested the co-integration approach during a relatively calm period, it would be very interesting to test it during a high volatile period. The main intention is not only to investigate whether a pairs portfolio is profitable or not but also to figure out how a portfolio containing co-integrated pairs behave against the market.

7.2 Portfolio data and analysis

For our analysis, we retrieved data from Yahoo Finance. More specifically, we retrieved all the closing prices from companies of the technology sector listed in the *New York Stock Exchange*, which is the largest stock exchange in the world by market capitalization of its listed companies. The reason we chose the corresponding sector of the *NYSE* is because this sector showed resilience despite the down turn of the market and recovered quite fast if we can say it.

In order to construct the pairs portfolio, we need to identify over the formation period companies that pass the co-integration test. The methodology is the same as we did in a previous chapter. However, applying statistical tests on a big number of assets, we expect to reject by chance the true null hypothesis. This is called in statistics Type I error. For the sake of simplicity we ignore this error. The formation period starts after the financial crisis of 2008. The reason we didn't retrieve data earlier is because old data might not reflect the current market conditions any more. The testing period starts in 2018 and lasts till late July 2020. So, as a first step, our algorithm creates a list of all the co-integrated companies within the technological sector. After that, we need to create a portfolio.

The initial step of constructing the portfolio is to decide the number of assets to add it. According to many studies, the diversification affect stops existing after adding more than 15 assets. This can be also confirmed in the below graph. We calculated the standard deviation of the logarithmic returns for various portfolios containing pairs. Our observation is that portfolios which contain more than 10 assets do not have significant difference in terms of the risk. Hence, our algorithm creates different portfolios which have the same weight. Also, the choice of the pairs is done randomly by the algorithm.

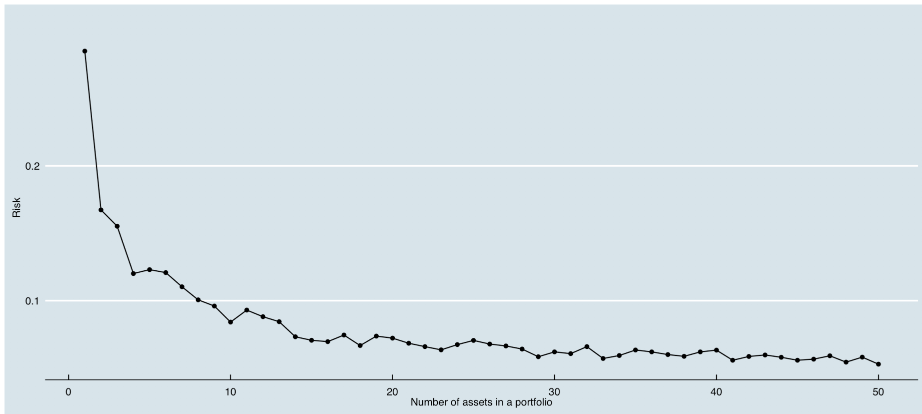

Figure 7.1: In the x-axis is presented the number of assets a portfolio contains. The y-axis shows the risk of each portfolio

According to figure [7.1], we decided to add 10 assets to each portfolio and to test it during the test period. The number of portfolios amounts to 100. For each portfolio we calculated the annual return and the volatility. What is more, we measure the beta parameter. As it has already been mentioned in previous chapters the beta is as follows:

$$\beta = \frac{\mathrm{Cov}(R_p, R_M)}{\mathrm{Var}(R_M)}, \tag{7.1}$$

where $\mathrm{Cov}(R_p, R_M)$ is the covariance of the portfolio returns and the market returns. The $\mathrm{Var}(R_M)$ term refers to the variance of the market returns.

Figure [7.2] shows that the majority of the portfolios do not generate any alpha. This essentially means that only a few of them outperform the market.

In figure [7.3] one can see the volatility of each portfolio relative to the market. As we might expected the volatility of all the portfolios is far below the market. The annual volatility of the market is close to 25%. However, the fact that market volatility is higher than each of the portfolios has not surprised us at all. The main reason of constructing a portfolio is to mitigate the idiosyncratic risk. However, the systematic risk as we know can not be eliminated.

It can be clearly seen in figure [7.4] that the beta of each portfolio is close to zero. At this point we can support the fact that the pairs trading is considered to be a beta neutral strategy.

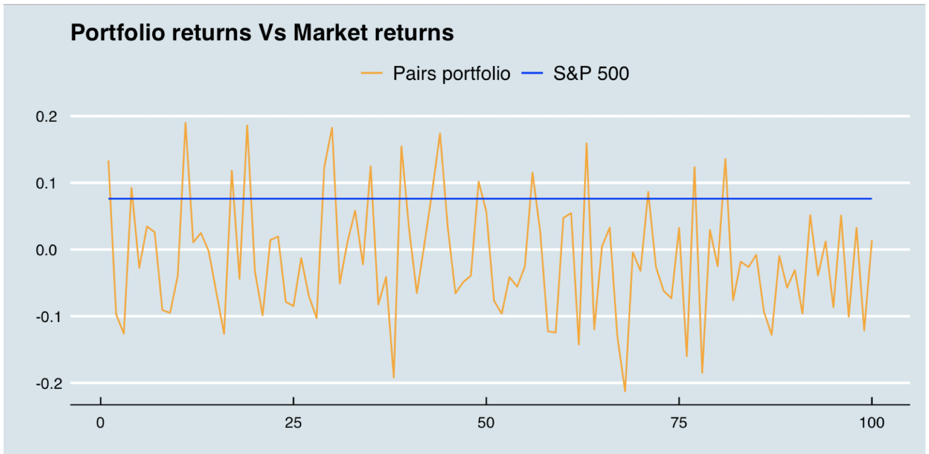

Figure 7.2: The y-axis shows the annul return in decimal points of each portfolio (orange line) and of *S&P 500* (blue). The x-axis represents the number of different portfolios

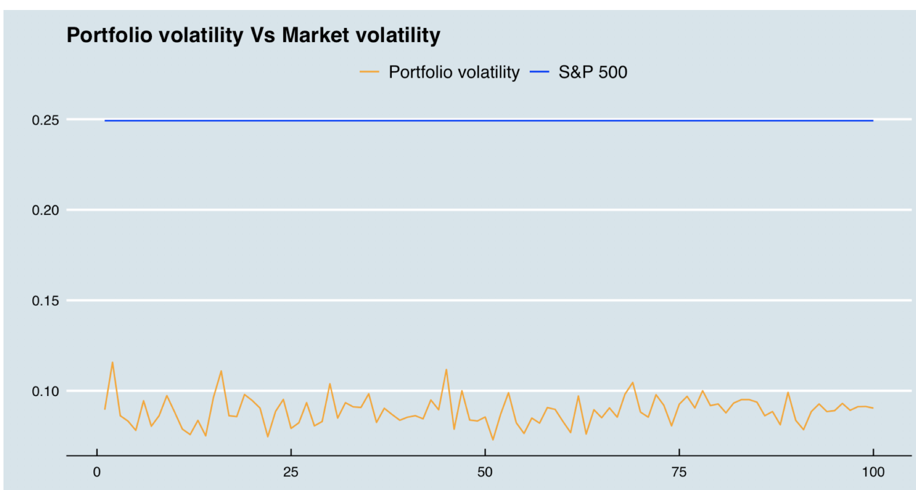

Figure 7.3: The y-axis shows the risk of each portfolio. The blue line shows the volatility of the *S&P 500*. The x-axis represents the number of different portfolios, while the orange line shows the volatility of each pair portfolio.

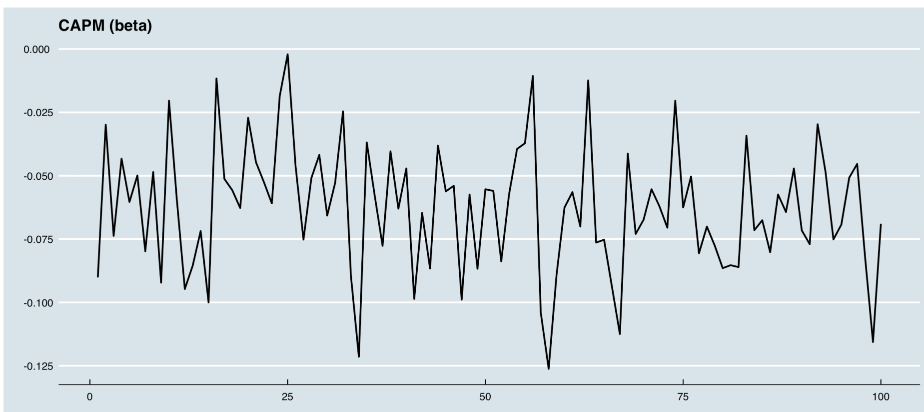

Figure 7.4: The y-axis shows the beta of each portfolio. The x-axis represents the number of different portfolios

Chapter 8

Conclusion

The final chapter deals with an overall conclusion of the study. It also provides some further ideas regarding a possible future extension of the development of the pairs trading strategy.

In this study, we tried to conduct a review of the mathematical aspects of pairs trading strategy based on the co-integration approach. Furthermore, we analysed through an empirical analysis, using real life data, the profitability of this strategy relying on different cross-validation techniques. What is more, risk measures were calculated as it is the most crucial part for each investment decision we take. In the last chapter, we took the opportunity to construct a portfolio of pairs. The main intention was to answer the question whether a portfolio of pairs is neutral relative to the whole market during a high volatile period.

However, we believe that the question regarding the performance of this method is a controversial issue in the financial industry. Many professionals believe that due to the high frequency trading, any co-movement opportunity is being eliminated in a zero time. On the other hand, not only used many studies the topic for research purposes but also successful financial institutions rely on mean revering strategies.

There is no doubt, although over the past decades pairs trading has become a very popular research topic, there is still room for carrying out further research on it. That would be very interesting to investigate the set up of a more profitable portfolio of pairs. Our idea was based on the choice of an equal weight portfolio containing pairs of the same sector. Therefore, we would also like to see a portfolio with pairs from different sectors. Another point is the parametrisation of the optimal look-back period. In our study we set up a fix 21 days looking back period. This essentially means that the algorithm takes 21 days back in history in order to adjust the new parameters taking from the regression. Our idea is to find a simple way to choose an optimal period based on the frequency that it takes to the spread to convert to its long term equilibrium.

References

[1] Ganapathy Vidyamurthy, (2004). *Pairs Trading Quantitative Methods and Analysis*. John Wiley Sons, Inc., Hoboken, New Jersey.

[2] Andrew Pole, (2007). *Statistical Arbitrage Algorithmic Trading Insights and Techniques*. John Wiley Sons, Inc., Hoboken, New Jersey.

[3] Douglas S. Ehrman, (2006). *The Handbook of Pairs Trading Strategies Using Equities, Options, and Futures*. John Wiley Sons, Inc., Hoboken, New Jersey.

[4] Ernest P. Chan, (2013). *Algorithmic Trading Winning Strategies and Their Rationale*. John Wiley Sons, Inc., Hoboken, New Jersey.

[5] Robert H. Shumway and David S. Stoffer, (2011). *Time Series Analysis and Its Applications*. Springer Science+Business Media, LLC.

[6] Paul Wilmott, (2009). *Frequently Asked Questions In Quantitative Finance*. John Wiley Sons Ltd, The Atrium, Southern Gate, Chichester, West Sussex, P019 8SQ, United Kingdom, 71-73.

[7] Carol Alexander, (2001). *Market Models: A Guide to Financial Data Analysis*. John Wiley Sons Ltd, The Atrium, Southern Gate, Chichester, West Sussex, P019 8SQ, United Kingdom.

[8] Han Xu, (2017). *High Frequency Statistical Arbitrage with Kalman Filter and Markov Chain Monte Carlo*. Master's thesis, Waterloo, Ontario, Canada.

[9] E. E. Holmes, M. D. Scheuerell, and E. J. Ward, (2019). *Applied Time Series Analysis for Fisheries and Environmental Sciences*. University of Washington, lecture material.

[10] Harry Georgakopoulos, (2015). *Quantitative Trading with R*. Palgrave Macmillian® in the United States—a division of St. Martins Press LLC, 175 Fifth Avenue, New York, NY 10010.

[11] Edwin J. Elton, Martin J. Gruber, Stephen J. Brown, William N. Goetzmann, (2014). *Modern Portfolio Theory And Investment Analysis 9th edition*. John Wiley Sons, Inc.

[12] Tomas Björk (2009). *Arbitrage Theory in Continuous Time* Oxford University Press.

[13] Kevin J. Davey (2014). *Building Winning Algorithmic Trading Systems* John Wiley Sons, Inc., Hoboken, New Jersey.

[14] Ali Hirsa (2013). *Computational Methods in Finance* CRC Press Taylor & Francis Group 6000 Broken Sound Parkway NW, Suite 300.

[15] Supakorn Mudchanatongsuk, James A. Primbs and Wilfred Wong. *Optimal Pairs Trading: A Stochastic Control Approach* 2008 American Control Conference paper, Pages 1035-1038

[16] Avellaneda, M. and Lee, J. H. (2008) *Statistical Arbitrage in the U.S. Equities Market.*

[17] Montana, G. (2009) *Flexible Least Squares for Temporal Data Mining and Statistical Arbitrage. Expert Systems with Applications*

[18] Burgess, A.N. (2000). *Statistical Arbitrage Models of the FTSE 100. In: Abu-Mostafa, etal. , Computational Finance 99*, MIT Press, Cambridge

[19] Thomaidis, N.S. and Kondakis, N. (2006). *An Intelligent Statistical Arbitrage Trading System. In: Lecture Notes in Computer Science* , Springer-Verlag, Berlin.

[20] Do, B., Faff, R. and Hamza, K. (2006). *A New Approach to Modelling and Estimation for Pairs Trading.*

[21] G. S. Maddala, In-Moo Kim, (1998). *Unit Roots, co-integration, and Structural Change*, Cambridge University Press.

[22] Zapart, C. (2003). *Statistical Arbitrage Trading with Wavelets and Artificial Neural Networks*, IEEE International Conference on Computational Intelligence for Financial Engineering, Hong Kong.

[23] Robert Jarrow, Melvyn Teo, Yiu Kuen Tse, Mitch Warachka (2012). *An improved test for statistical arbitrage*, Journal of Financial Markets, Pages 47–80.

[24] Evan Gatev, William N. Goetzmann, K. Geert Rouwenhorst (2006). *Pairs Trading: Performance of a Relative Value Arbitrage Rule*, The Review of Financial Studies

[25] Robert F. Engle and C. W. J. Grange (1987). *Co-Integration and Error Correction: Representation, Estimation, and Testing*, Econometrica, p253.

[26] RJ Elliott, J Van Der Hoek, WP Malcolm (2005). *Pairs Trading. Quantitative Finance*, Vol. 5, Pages 271-276

[27] Markus Harlacher (2016). *co-integration Based Algorithmic Pairs Trading*, Doctoral dissertation, University of St. Gallen