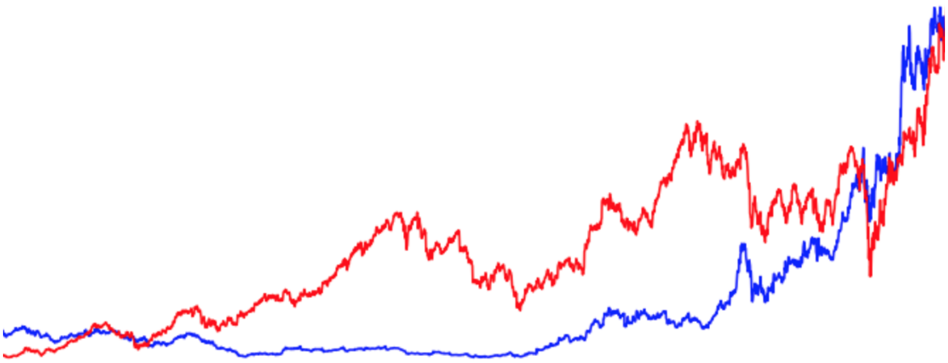


Developing short term algorithmic pairs trading strategies using time series modeling

Afonso Villaret Frischknecht

Thesis to obtain the Master of Science Degree in

Mechanical Engineering

Supervisors: Prof. João Miguel da Costa Sousa
Prof. Artur Jorge da Cunha Barreiros

Examination Committee

Chairperson: Prof. Duarte Pedro Mata de Oliveira Valério
Supervisor: Prof. Artur Jorge da Cunha Barreiros
Member of the Committee: Prof. Nalan Basturk

July 2021

Acknowledgments

I would like to thank my supervisors, Professors Artur Barreiros and João Sousa, for their never-ending availability, support and encouragement, throughout this work's development. They have most definitely played an active and decisive role in my learning process.

To my family, as they are my life's foundations, my confidants, advisors and best friends.

Last but not least, to my student colleagues, loyal and trusted friends, Guilherme Kano, José Correia and Luís Monteiro, for their always guaranteed help, guidance and friendship. They were, undeniably, responsible for assuring my academic success.

Resumo

Nesta dissertação procurámos aplicar conceitos relativos à modelação e identificação de séries temporais, na implementação de estratégias de negociação de activos. Focámos a nossa atenção em estudar estratégias de investimento que sejam independentes de condições económicas, e cujos retornos, consequentemente, sejam neutros ao comportamento do mercado. Mais especificamente, dedicámo-nos à negociação de pares, envolvendo a abertura simultânea de uma posição longa, e de uma posição curta, em cada um dos activos emparelhados.

Para que estratégias como esta sejam bem sucedidas, é necessário realizar uma análise cuidada e detalhada dos dados históricos das acções e pares negociados. Assim sendo, adotámos diversas metodologias de decisão baseadas no modelo CAPM e no exponente de Hurst. Para cada uma delas, analisámos o impacto, nos respectivos resultados, de diversos parâmetros indispensáveis à sua implementação. Com este objectivo em mente, recorremos ao desenvolvimento de algoritmos de negociação automatizada em Python. Em última análise, o método baseado no exponente de Hurst revelou os melhores resultados.

Todos os passos do nosso raciocínio e método serão detalhadamente expostos e discutidos.

Palavras-chave: Séries temporais, Negociação de Pares, estacionariedade, Cointegração, Hurst, CAPM.

Abstract

In this dissertation, we have applied concepts related to time series' modeling and identification, on the implementation of speculative investment strategies. We have focused our attention on studying strategies which are independent from economic conditions, and whose returns, as a consequence, are neutral to market behaviour. More specifically, we have engaged in exploring pairs trading strategies, which involve the simultaneous opening of both a long, and a short position, in each of the traded assets.

For strategies like these to be successful, it is necessary to carry out a careful and detailed analysis of the historical data from shares and pairs. With that in mind, we have adopted several decision methodologies based on the CAPM and on the Hurst Exponent. For each of them, we have analyzed the impact, on their respective results, of diverse indispensable parameters regarding their implementation. With this purpose, we have developed an in house automatic decision algorithm, resorting to Python. This algorithm has allowed us to perform dozens of simulations, and virtually trade hundreds of pairs across different trading periods. Ultimately, the Hurst Exponent based methodology revealed the best results.

All steps of our reasoning and method will be exposed and discussed in detail. The results of all the simulations, carried out throughout this work, will also be subject to rigorous scrutiny.

Keywords: Time Series, Pairs Trading, Stationarity, Cointegration, Hurst, CAPM.

Contents

List of Tables

List of Figures

Glossary

- **Long Buy**: strategy which involves the purchase of a stock or asset with the expectation of increasing its value.

- **Short Sell**: investment strategy by burrowing a stock from a brokerage and selling it immediately, with the expectation of returning it by buying it at a lower price.

- **S&P 500**: Stock market index that measures the performance of the 500 largest publicly traded companies in the United States.

- **Volume**: A given financial asset's number of traded shares, over a certain period of time.

- **Liquidity**: Refers to how easily can a stock or an asset be converted into cash.

- **Bearish Trend**: Downward trend in the stock's price.

- **Bullish Trend**: Upward trend in the stock's price.

- **Adjusted Close** Corrected stock's closing price after accounting any corporate action.

- **Broker**: Individual or firm that facilitates the contact between investors and security exchanges. Brokers provide those services and earn compensations in diverse forms (normally via commissions or fees).

Chapter 1

Introduction

1.1 Overview on pairs trading

People have long pursued profit by either buying, selling or simply trading resources and goods. This quest, soon led to the spread of speculative investment strategies strategies. Ever since the appearance of the first companies, investors have attempted to position themselves in ways that allow them to favourably explore price fluctuations of financial assets. Numerous theories, on why and how these fluctuations occur, have been developed and studied over time. Several factors, ranging from human conflict to natural catastrophes, can have severe economic impact, and play a decisive role, in determining a given asset's valuation.

Nowadays, one of the most widely adopted theories is known as Capital Asset Pricing Model (CAPM), and was first proposed by Sharpe [1]. This theoretical model, suggests that favourable economic conditions tend to favour the great majority of assets, while negative economic conditions tend to do the exact opposite. In fact, it states that the return on a given asset is largely influenced by the return on the market, during the considered period. This theory, attempts to describe asset returns as the sum of two separate components: the residual or non-systematic term, and the market or systematic term. The return on the asset is then modeled as the return on the market, times a ratio factor known as Beta, plus a residual return, whose value is independent from economic conditions and market behaviour.

In reality, when economic conditions seem positive, investors expect the price of securities to rise, in a generalized way. As a consequence, the purchase of assets is often viewed as the most appealing option. On the other hand, if an adverse economic scenario exists, people usually predict a downfall in prices and valuations. When this happens, investors tend to sell assets they own, or even, in some cases, resort to short selling. Nonetheless, these strategies were still vulnerable to rapid economic changes, and unforeseen market alterations.

Armed with these notions, investors and academics soon began searching for ways that would allow them to invest in assets without being dependent on market fluctuations. These are called Market Neutral Strategies and the most commonly used one is Pairs trading. Pairs trading is a strategy in which a portfolio composed by two assets is created. The objective is for the return on this pair of assets to behave independently from economic conditions. However, as we shall see, this is not always possible.

Such methodologies, are often referred to as statistical arbitrage pairs trading strategies. These speculative procedures, rely on historical data regarding financial products. This data, is studied and analyzed by investors, in an attempt to detect future statistical opportunities.

The basic principle, for any pairs trading strategy, is to go long on one of the assets, and short sell the other in a predetermined weight ratio. By doing this, the investor no longer relies on the particular valuation of any of the securities but, instead, depends sole on their relative valuation. In other words, the spread between both price series becomes the investment product.

For this strategy to work, it is important for both assets to show a long lasting price equilibrium. When this happens, and the resulting spread, between the two price series, reveals itself as a stationary mean-reverting time series, the asset pair is said to be cointegratable. The purpose of the investor in these cases, is to explore relative price deviations. Every time the spread deviates substantially (either by excess or default) from the mean, investors enter the market in a position that will allow them to profit if mean reversion occurs. When this happens, the positions are withdrawn, investors exit the market, and the subsequent profits or losses are taken.

As simple as this may seem, over the years, numerous literature has consistently proved the profitability of pairs trading strategies. Some authors and academics have stated that, when properly implemented, such strategies can even profit from negative market conditions such as the recent 2007-09 global financial crisis [2].

This being said, one should ask what the conditions for this success are and what risks exist. Several authors, such as Vidyamurthy [3] and Gatev et al. [4], have studied these strategies and attempted to develop practical guidelines to effectively apply and implement them. By doing so, one of the first questions that arises is how stocks can be chosen to form the pair. By intuition, the reader may be tempted to answer "test all pair possibilities, and then test for cointegrability". However, in an universe composed of thousands of stocks, hundreds of millions of pairs can be formed, and applying cointegration tests on all of them is not a valid option. The answer to this question has been widely debated, and several approaches have been presented and tested without consensual results.

Once securities are chosen, pairs are formed, and the resulting spread between the assets is computed. Subsequently, a proper study and analysis of these spread series is required. Pairs whose spread series depict the desired behaviour must be selected for trading, while pairs with non-stationary spread series must be discarded.

Huck and Afawubo [5] developed studies where comparisons between different pair filtering methodologies where performed. Tests were ran for Distance, Stationarity and cointegration designed algorithms. The cointegration based strategies ultimately revealed the best results. Although many other strategies, based on different approaches, exist, such as Combined forecasts and Multi-Criteria Decision Methods (MCDM) [6], for the purpose of this dissertation, we will only consider the cointegration based ones.

Through the course of this work, all of these concepts will be applied in designing and implementing, an algorithmic Statistical Arbitrage Pairs Trading strategy. Algorithms for selecting and matching stocks, and for filtering the resulting pairs, will be developed and analysed. In addition, filtered pairs will automatically undergo a trading period in which the strategy will be put to the test. Several hundred simulations will be performed, and dozens of pairs tested across several different trading sessions. The performance of all these simulations will be closely monitored and their results discussed.

1.2 Motivation

It is easy to understand that the world of finance, and in particular, the world of trading, is a never ending environment for studies, improvements and optimizations. Despite the extensive discussion pairs trading has merited over the years, several queries regarding its implementation are still pending and open for further investigation.

One specific area, where authors struggle to provide unanimous answers, is how stocks can be selected and the subsequent pairs filtered for testing. The ability of an investor to, a priori, determine which assets are best candidates for cointegration, the time period one should account for when studying the stocks, and for how long one should keep the strategy running unattended, are just some examples. The objective of this work is not, by any means, to present a guarantee way for profit, but instead, an attempt to formalize guidelines and principles that can improve and simplify the implementation of Statistical Arbitrage Pairs Trading strategies.

Several statistic tests will be performed on historical price data from numerous stocks quoted in the US Stock Market. Several assumptions and simplifications will be adopted in comparison to a real investment environment. However, the main goal is for this thesis to provide a valid, segmented approach to the problem. Tests are also expected to present helpful conclusions regarding future implementations, in more realistic settings.

Stock markets are ever changing dynamic systems, whose prediction is a complicated, and effort demanding, procedure. As a Mechanical Engineer to be, I consider important to demonstrate my ability to properly examine and scrutinize, any kind of time variant, dynamic process.

In this dissertation, I will present an engineering perspective on how time series can be analyzed and modeled. While doing so, I'll also study several, indispensable, parameters for creating, designing and testing, any historical data based, trading strategy. As a student of a systems area masters degree, I truly believe this can be an ideal opportunity for applying concepts and techniques acquired during the course of my engineering studies. These tools will range from statistical instruments, to system identification methods and data processing procedures. Software development in a new language, as well as the acquirement of knowledge in a different field, are also major motivators.

1.3 Contributions

We set out with the objective of developing functional and applicable procedures, for implementing short term statistical arbitrage pairs trading strategies. With that in mind, we have adopted three different asset selection methodologies (all based on the assets' price series), and developed algorithms, to test and compare them. In addition, several important, time varying parameters, were studied and analyzed, for each selection setting. All things considered, there are several, worth noting contributions, to retrieve from this work.

We provided a detailed description of the three different stock selection methodologies adopted in this dissertation. Two were based on the CAPM, and formed pairs by combining assets with either high or close to unit Beta values. In the third one, stocks with low Hurst Exponent values were selected and matched. It became clear that, not only are these methodologies vital for a practical implementation of the strategies, but they are also key in ensuring the quality of their results.

When we compare the global performance of the three selection methods, we can observe an undeniable improvement of results, for the low Hurst Exponent criterion. In fact, this method revealed the greatest number of positive return strategies, as well as the best Hit and Sharpe ratio results. Furthermore, it became clear that pairs constructed by combining stocks with low Hurst Exponent values, are much more likely to be cointegratable, than those derived from joining together stocks with similar Beta values.

Furthermore, we have examined the results obtained for different training and trading length configurations. Once again, valuable deductions could be made, as shorter term strategies ultimately revealed better performance indicators.

1.4 Thesis Outline

In Chapter 2, we have provided some vital background concepts and methodologies used in implementing statistical arbitrage pairs trading strategies. We have also presented some indispensable time series models as well as explained the need for an unit root test. In addition, the detailed procedure used for this test is also bestowed in this segment.

In Chapter 3, we carefully explained the entire decision making process behind our strategy. This process will range from data processing mechanisms to the opening and closing of buy and sell positions.

Throughout Chapter 4 several performance monitoring tools were presented and examined. These tools are key to correctly understand the strategies performance, as well as to properly quantify the magnitude of its success or failure.

The results from all these performance indicators is shown in Chapter 5. In this segment it is possible to observe the obtained results from all selection procedures and for different training and trading length settings.

Finally, in Chapter 6, we have discussed and debated the conclusions of our work. We have also provided valuable verdicts and guidelines that can be helpful to a real life implementation of such strategies.

Chapter 2

Pairs trading, time series and unit root testing

The implementation of statistical arbitrage pairs trading strategies, relies on a thorough analysis of price, return and spread, time series. This meticulous examination, has the purpose of correctly identifying the model which best describes those series.

In this chapter, we will present some vital, background concepts, on statistical arbitrage pairs trading, as well as a detailed revision of essential time series models. We will also explain the need for an unit root test, along with a detailed description of its methodology.

2.1 Cointegration Based Statistical Arbitrage Pairs Trading

Statistical Arbitrage Pairs Trading refers to a set of strategies in which investors attempt to explore price disequilibriums between cointegrated assets. These are market neutral procedures, since they involve the simultaneous opening of both long and short positions in a given pair of financial products. The goal is to explore possible relative mispricing regarding the pair. As a consequence, when properly implemented, investors are not dependent on market fluctuations, but instead, in order to profit, they rely sole on the relative pricing of the two assets. When the spread of the two cointegrated prices diverges by more than a certain threshold value, the relatively overvalued asset is short sold and the relatively undervalued asset is long bought. Upon convergence of the prices, the position in unwind and profit, or loss, is taken due to the variation of the spread series.

Pairs Trading was first introduced in the 1980s by American investment banks operating in Wall Street. These banks hired teams of technical analysts and encouraged them to develop speculative strategies based on asset prices' historical data. Researchers would analyse this data and pursue pairs of financial products whose prices exhibited a long lasting equilibrium. The main assumption, behind any statistical arbitrage pairs trading strategy, is that the historical behaviour will exhibit itself once again in the future. Therefore, since its profits rely on future confirmation of this hypothesis, performing a detailed and thorough data analysis is vital for success.

One other major assumption, necessary to fully understand such strategies, is related with market efficiency. Market efficiency is a principle which stands for the market's ability to properly and effectively, detect and correct relative mispricings. In other words, financial actors are expected to recognize market inefficiencies and promptly correct them. For example, if two stocks, representing similar value companies, reveal a price discrepancy, investors should notice this and invest their money accordingly. By doing so, and assuming that many other individuals will also recognize this deviation, the market will once again force the prices of the two stocks to an equilibrium point. This behaviour is often referred to as Mean Reversion, and it is imperative for the profitability of any statistical arbitrage based strategy.

As soon as these principles became public, several researchers and academics dedicated themselves to extensively study and analyse them. Authors quickly began searching for methods and procedures that would allow them to formulate effective, practical guidelines, for the implementation of such strategies. As a consequence, numerous literature on Pairs Trading has been developed ever since.

The way financial products can be chosen, paired and filtered is indispensable for properly applying this methodology. Quite obviously, the first implementation step faced by investigators is related to asset selection. Despite eventual temptations to combine thousands of stocks, proceeding to filter all the resulting pairs is a virtually impossible task. As a consequence, procedures regarding stock selection are crucial for trading pairs. Although several approaches have been adopted, academics haven't yet reached consensus.

Vidyamurthy [3] attempts to solve the problem by formalizing a notion of similarity between stocks. This distance based notion, defines an asset's return as a sum of two components: common factor returns (which result from different risk factors, and are common to all assets), and specific returns, unique to each stock. The absolute value of the common factor correlation between the two assets is computed. This correlation is the same as the correlation between the two innovation sequences of each stock. The specific variance contribution is not considered. A candidate list is then formed. If the common factor correlation is either +1 or -1 (meaning the factor exposure vectors are perfectly aligned) and the specific return series are stationary, than the conditions for cointegration are satisfied. These ideal cointegration conditions are difficult to achieve in practice. It is, however, possible to implement the strategy on stocks which deviate slightly from them.

Gatev et al. [4], on the other hand, follow a different methodology. The authors start by constructing an index of cumulative total returns for each stock, and then choose a suitable match by finding the asset that minimizes the sum of the squared deviations, between the two normalized price series. The author then advances to prove that stocks from the same industry are more likely to be cointegratable, and tend to reveal better results.

Several other approaches have been used and it is interesting to note that, most of them, attempt to find stocks whose prices move together. Numerous literature also suggests that statistical arbitrage pairs trading strategies tend to offer better results when implemented in stocks with high levels of liquidity.

On this dissertation, selection of assets will be based on two main methodologies: stocks which present similar behaviour patterns in relation to the market (Section 3.2.1), and stocks with low Hurst Exponent values (Section 3.2.2).

Once pairs are formed, it becomes imperative to perform a detailed spread series analysis, with the purpose of properly filtering such portfolios. Again, countless methodologies and approaches regarding pair filtering have been developed throughout the years.

One particularly important article, to understand the approach used in this dissertation, is Huck and Afawubo [5]. In this work, the authors start to form pairs by joining together stocks whose returns during the selection period are similar (differ by less than 10%). Then, they run a series of tests in order to explore the performance of a Pairs Trading system based on various pair filtering methods. More specifically, investigators compare the results from three of the most commonly used methodologies:

- the minimum distance approach,
- the stationarity of the price ratio and
- the cointegration between stock prices

The minimum distance method has been vastly adopted in the finance world since Gatev et al. [4] (2006). Do and Faff [2] also developed important studies regarding this approach's implementation. In this simple methodology, the first step is to normalize stock prices. Afterwards, pairs are formed by finding a suitable matching stock. The matching criterion is the minimization of the sum of the squared differences between the daily normalized prices, during the formation period. The pairs with the lowest sum are then chosen to undergo a trading period in which the relative pricing will be explored.

As for the stationarity of the price ratio approach, once again assets are primarily selected based on the similarity of returns. Pairs are then formed and the price ratio among them is tested for stationarity. Note that in order for the strategy to succeed, test results must reveal a long lasting equilibrium, with constant mean and variance. When this verifies, the pair is subjected to a trading period in which substantial deviations from the price ratio are interpreted as trading opportunities.

The third and last filtering method is central to this dissertation's scope. Cointegration was first proposed in 1987, by Nobel laureates Engle and Granger [7]. Cointegration tests allow for identification of time periods in which two, or more, non-stationary time series are merged, originating a series that is unable to deviate from equilibrium in the long term. More specifically, cointegration occurs when two non-stationary variables, y_t and x_t can be expressed as a stationary process u_t, such as

$$u_t = y_t - ax_t \tag{2.1}$$

where a is a weight constant for the non-stationary variable x_t.

Over the years, several different cointegration tests were developed and published by numerous academics. For the purpose of this work, the Augmented Dickey Fuller test (ADF), merits special emphasis. The ADF test was developed by Dickey and Fuller [8] in 1979 and has been extensively used in time series modeling ever since. It will be further addressed in 2.3. In fact, the conclusions of Huck and Afawubo [5] suggest that cointegration based systems reveal far better results than any of the other two methodologies. While the minimum distance and the stationarity of the price ratio approaches displayed results between 0.27%-0.33% and 0.36%-0.48%, respectively, cointegration based systems accomplished returns between 2.08% and 5.86% per month, over a period of more than 10 years.

In our case, two non-stationary time series, representing the prices of the two assets which form the pair, are integrated together in a pre-determined weight ratio. The resulting series, representative of the spread between the two stocks, is then tested for cointegrability. With that purpose, an ADF test is performed on the time series. If the result of the test reveals stationarity of the spread series, the pair will pass the filtering stage of the strategy and, therefore, undergo a trading period. If not, the pair is discarded and does not get traded.

Pairs that successfully pass the filtering stage, are tested, and traded, during a certain time period. During this period, the spread of the pair will be closely monitored in search of substantial deviations from the mean. The threshold used for defining the opening of a long-short position is based on the standard deviation of the spread series. If the spread shortens bellow the threshold value, a long position will be assumed on the higher value stock and a short position will be assumed on the lower value stock. If the spread widens above the threshold, inverse positions will be adopted. When the spread reverts to its mean value, or if the time period for trading ends, the positions are withdrawn.

2.2 Time Series

As previously stated, the design and implementation of Statistical Arbitrage Pairs Trading strategies, banks on a thorough analysis of the time series which represent asset prices and asset returns. The proper identification and modeling of such series is key for success. Therefore, the following section is dedicated to presenting vital models and concepts, regarding the development of time series based trading algorithms. The reader is assumed to have basic statistic knowledge.

2.2.1 Stationarity of a Time Series

As prior explained, the stationarity of a given pair's spread series is a necessary condition for the pair to be used in the strategy. As a consequence, we will begin our analysis by defining stationarity of a time series.

A time series is said to be Strictly Stationary if the marginal distribution of its value at time instance t, $(p(y_t))$, is the same as at any other point in time. This condition can be written as

$$p(y_t) = p(y_{t+k}) \tag{2.2}$$

where $t \geq 1$ and k is a constant integer. This implies that the mean, variance and covariance of the y_t series, is time invariant.

A time series can also be said to be Weakly Stationary, or Covariance Stationary. In these cases, its expected value, variance and covariance, must meet the following conditions:

- $E(y_1) = E(y_2) = ... = E(y_t) = \mu$
- $V(y_1) = V(y_2) = ... = V(y_t) = \gamma_0$
- $COV(y_1, y_{1+k}) = COV(y_2, y_{2+k}) = ... = COV(y_t, y_{t+k}) = \gamma_k$

where μ and γ_0 are constants, and γ_k depends only on lag k.

As for non-stationary time series, they can be classified as such either by presenting non-constant means or non-constant variances. When the series reveal a non-constant mean they can be divided into two main trend groups:

- Deterministic Trends
- Stochastic Trends

Deterministic Trends occur when a series is trending because it presents an explicit relation with time. These models fall out of this dissertation's scope since they do not allow for stock price representation.

As for Stochastic Trends, they occur in series whose current value is dependent on previous series' values. A common example of such a trend is an AR(1) model (Equation (2.12)). This model will be later explained in section 2.2.5.

There are, however, non-stationary time series that can be made stationary by differencing. When this can be accomplished by differencing the series one time, they are said to be Integrated of order one and are denoted as $I(1)$. Analogously, if a given non-stationary time series can be made stationary after differencing it d times, it is said to be Integrated of order d, and is denoted as $I(d)$.

Series which are already stationary before being differenced, can also be referred to as Integrated of order zero, and be denoted as $I(0)$.

Having defined stationarity, it is now indispensable to introduce the concept of auto correlation.

2.2.2 Auto Correlation Function (ACF)

The Autocorrelation Function (ACF), measures the degree of relationship between a given time series and a previous version of itself, lagged τ time intervals apart, and it is expressed by

$$\hat{\rho}\,(\tau) = \frac{\frac{1}{T} \sum\limits_{t=\tau+1}^{T} [y_t - \overline{y}] - [y_{t-\tau} - \overline{y}]}{\frac{1}{T} \sum\limits_{t=1}^{T}} \tag{2.3}$$

where T is the total number of time intervals, τ the time lag to be considered and $\overline{y}$ the mean average of the time series y. The higher the value of $\hat{\rho}\,(\tau)$, the higher the correlation between the value of the time series at instance t (y_t) and its value, τ time lags before ($y_{t-\tau}$).

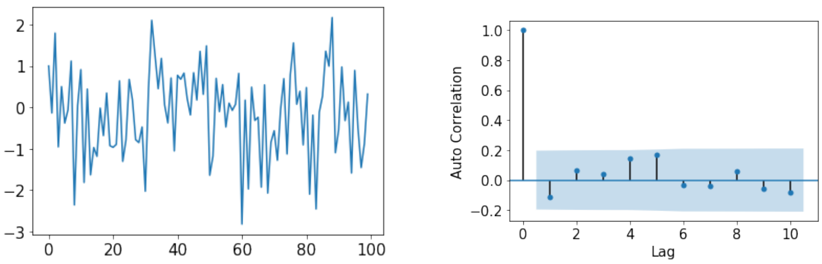
Armed with these notions, we can proceed to examine some crucial models for describing time series. An accurate identification of the best fitting model is needed for the strategy's success.

2.2.3 White noise

White noise (Fig.2.1 (a)), is a purely random time series model, whose values are sampled, at each time instance, from uncorrelated Gaussian distributions. These samples are denoted as ϵ_t. The mean, μ, and the variance, σ^2, of this distribution, are fixed and do not change over time. The value of the time series at instant t, y_t, can be expressed as

$$y_t = \epsilon_t \tag{2.4}$$

Each value is uncorrelated to the rest of the time series, as depicted by the ACF function of the model, presented in Fig. 2.1 (b). This is due to the fact that they are drawn from uncorrelated normal distributions, with the same mean and variance.



(a) Example of white noise with $\mu = 0$ and $\sigma^2 = 1$.

(b) White noise's auto-correlation plot.

Figure 2.1: White noise (left) and respective ACF (right).

As you can see, the plot of such series over time, does not reveal any clear pattern. Forecasting or predicting white noise series is, therefore, impossible. However, and more importantly, white noise models represent stationary, mean reverting time series.

2.2.4 Moving Average (MA)

A Moving Average model of order q, denoted as MA(q), represents a time series where its value, at the current time instance t, z_t, is the sum of the q previous white noise realizations, ϵ_{t-q}, plus the mean of the series, plus an innovation term. Thus, the value of the series at time step t, only depends on random error terms, which follow a white noise process, with $\mu = 0$ and $var = \sigma^2$. A generic MA(q) model is represented as

$$z_t = \overline{z} + \epsilon_t + \beta_1\epsilon_{t-1} + \beta_2\epsilon_{t-2} + ... + \beta_q\epsilon_{t-q} \tag{2.5}$$

where $\overline{z}$ is the mean of the series, β_i are the model's parameters, and ϵ_t represents the white noise realisation at t^{th} time step. It is also possible to rewrite equation (2.5) in a manner that simplifies our analyses.

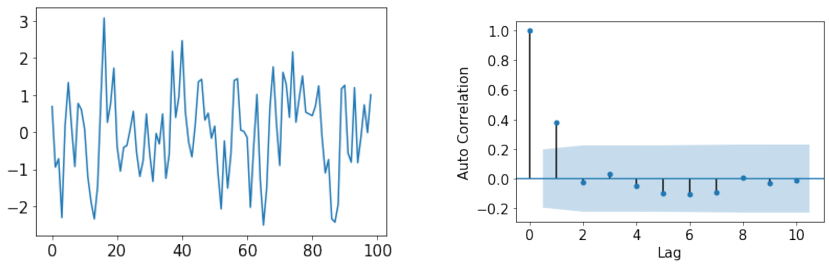
$$y_t = \epsilon_t + \beta_1\epsilon_{t-1} + \beta_2\epsilon_{t-2} + ... + \beta_q\epsilon_{t-q} \tag{2.6}$$

where the variable transformation, $y_t = z_t - \bar{z}$, allows us to interpret the process, as being sole dependent on the past and present white noise realizations.

Assuming a Moving Average model of order 1, equation MA(1), (2.5) can be rewritten as

$$y_t = \epsilon_t + \beta\epsilon_{t-1} \tag{2.7}$$

Fig. 2.2 (a) represents an example of a MA(1) time series, with the corresponding ACF shown in Fig. 2.2 (b), which suggests a correlation between the current time series value, y_t, and the previous time step, y_{t-1} due to the common white noise realisation, ϵ_{t-1}.



(a) MA(1)

(b) MA(1)'s auto-correlation function.

Figure 2.2: MA(1) (left) and respective ACF (right).

As you can see, it is also obvious that in the second time lag, the MA(1) model's ACF, decays abruptly to zero. If we were to consider a MA(2) model,

$$y_t = \epsilon_t + \beta_1\epsilon_{t-1} + \beta_2\epsilon_{t-2} \tag{2.8}$$

the current value of the series would be a function of the two past white noise realizations plus the innovation term. As a consequence, the model's ACF would only decay to zero at the third time lag. If we continuously repeat this process, we arrive at a MA(∞) model, described as

$$z_t = \mu + \epsilon_t + \sum_{j=1}^{\infty} \beta_j z_{t-j} \tag{2.9}$$

This particular model can be a representation of an Autoregressive model of order 1 (Equation 2.12).

2.2.5 Autoregressive (AR)

In a basic linear regression, the dependent variable is modelled as a linear function of the independent variable plus a random error term

$$y_i = \beta_0 + \beta_i x_i + \epsilon_t \tag{2.10}$$

An Autoregressive time series of order p, denoted as AR(p), represents a time series whose value at time step t, y_t, is a linear combination of its p previous values plus a realisation of white noise at the time step t. This model can be written as

$$y_t = \gamma_0 + \gamma_1 y_{t-1} + \gamma_2 y_{t-2} + ... + \gamma_p y_{t-p} + \epsilon_t \tag{2.11}$$

where γ_i are the model's parameters and y_{t-i} is the value of the series, at time instance $t-i$. For a AR(1) time series, equation (2.11) can be written as

$$y_t = \gamma_0 + \gamma_1 y_{t-1} + \epsilon_t \tag{2.12}$$

For this model, the expected value of the time series at instant t, $\overset{*}{y}_t$, will be given by

$$\overset{*}{y}_t = \gamma_0 + \gamma_1 y_{t-1} \tag{2.13}$$

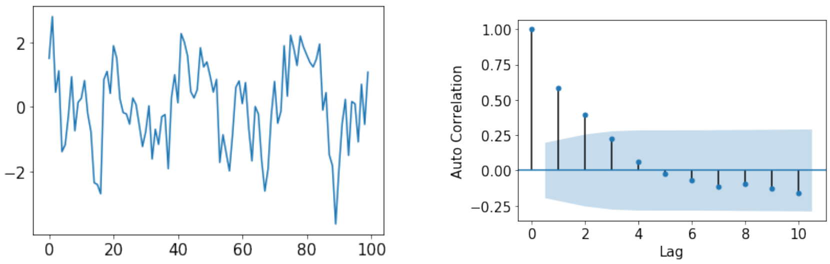
In this case, the variance of the expected value would be equal to the variance of ϵ_t, which is the same as the variance of the white noise series used to construct the model.

As previously stated, AR(1) models can be manipulated to arrive at a MA(∞) process. If we consider the value of the AR(1) model, y_{t-1}, at the previous time instance, $t-1$, we have

$$y_{t-1} = \gamma_0' + \gamma_2 y_{t-2} + \epsilon_{t-1} \tag{2.14}$$

The combination of equations (2.12) and (2.14), yields that we can continuously replace the terms regarding the previous values of the series, by others, representing the white noise realizations at each time step. By doing so, it is possible to transform the model into a MA(∞) process.

Fig. 2.3 (a) describes a generic AR(1) time series model. Its corresponding ACF is shown in Fig. 2.3 (b). In autoregressive models, the ACF exhibits a slow decay as a direct consequence of its linear dependence on past time series values, as evidenced by Fig. 2.3 (b).



(a) AR(1) (b) AR(1)'s auto-correlation plot.

Figure 2.3: AR(1) (left) and respective ACF (right).

2.2.6 ARMA

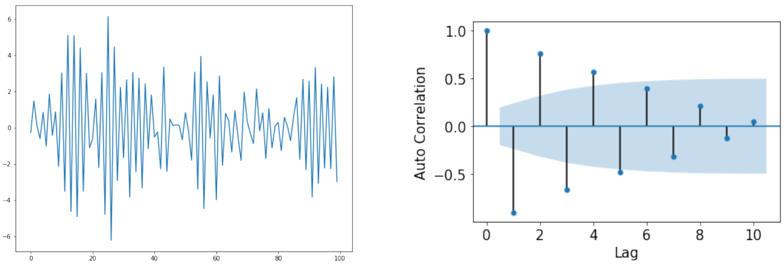
There are situations where time series may be represented as a mix of both AR(p) and MA(q) models. These models are called ARMA(p,q). They depend on their p past values, as well as on the last q white noise realizations. ARMA(p,q) models can be expressed as

$$
\begin{aligned}
y_t = & [\gamma_0 + \gamma_1 y_{t-1} + \gamma_2 y_{t-2} + ... + \gamma_p y_{t-p}] + \\
& + [\epsilon_t + \beta_1 \epsilon_{t-1} + \beta_2 \epsilon_{t-2} + ... + \beta_q \epsilon_{t-q}]
\end{aligned} \tag{2.15}
$$

or, using a different terminology

$$
y_t = \gamma_0 + \sum_{j=1}^{p} \gamma_j y_{t-j} + \sum_{j=1}^{q} \theta_j \epsilon_{t-j} + \epsilon_t \tag{2.16}
$$

Fig. 2.4 (a) describes a generic ARMA(1) time series model. Its corresponding ACF is shown in Fig. 2.4 (b). In autoregressive moving average models, the ACF exhibits a slow decay as a direct consequence of its linear dependence on past time series values, as evidenced by Fig. 2.4 (b).



(a) ARMA(1,1)

(b) ARMA(1,1)'s auto-correlation function.

Figure 2.4: ARMA(1,1) (left) and respective ACF (right).

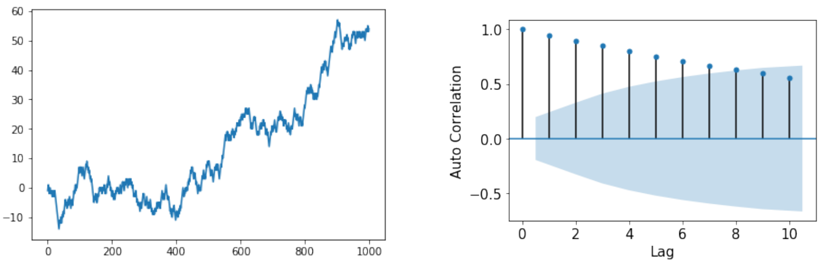
2.2.7 Random walk

A Random Walk process is a specific case of an AR model that is often used to model phenomena such as stock prices' fluctuations, and which consists in a AR(1) series with $\gamma_0 = 0$ and $\gamma_1 = 1$. According to the definition stated in equation (2.12), a Random Walk is defined by

$$
y_t = \epsilon_t + y_{t-1} \tag{2.17}
$$

which states that the value of the series at time instance t, y_t, is the sum of the previous value of the series, y_{t-1}, and the current realisation of the white noise at instance t, ϵ_t, commonly known as the innovation term.

As in other models, the expected value for the time series at instance t, $\overset{*}{y}_t$, is the previous value of the series y_{t-1}. However, and although this property can have an effect on the ACF's decay rate (Fig. 2.5), since the white noise term can not be predicted, all Random Walk processes are, by nature, unpredictable. One other important characteristic regarding such models is that they can revel trends. In fact, all Random Walk processes are non-stationary.



(a) Random Walk.

(b) Random Walk's auto-correlation function.

Figure 2.5: Random Walk (left) and respective ACF (right).

2.3 Unit Root Testing

The time series models introduced in the previous sections are vital, since they are to be used in representing the spread series of the studied pairs. Nonetheless, we still lack a procedure that permits us to correctly identify the most accurate representation of such series.

Cointegration between two time-series can be evaluated according to the definition presented on Equation (2.1). As for stationarity, it occurs when a shift in time does not provoke an alteration in the parameters of its distribution. Unit roots are one cause for non-stationarity. Unit root tests, as the name suggests, examine the existence of unit roots in time series. With this purpose, several tests have been developed over the years, such as the Elliott–Rothenberg–Stock test [9] and the Schmidt–Phillips Test [10]. Through this dissertation's scope, we will only consider the Augmented Dickey-Fuller test (ADF).

2.3.1 ADF Test

The Dickey-Fuller (DF) test, was first introduced by Dickey and Fuller [8]. Phillips and Perron [11] have successfully applied this statistical method to test a null hypothesis regarding the existence of an unit root in a time series and, therefore, inspecting its stationarity behaviour.

In [8], the authors perform the test upon a first-order regressive model, AR(1):

$$y_t = \phi y_{t-1} + \epsilon_t \tag{2.18}$$

The implication of a model of this type is that the best prediction for the next value of the series is its own previous value. Nevertheless, an AR(1) model can alter its behaviour as a consequence of the value of the constant parameter ϕ. As previously explained, if $\phi = 1$ the model becomes a Random Walk process. On the other hand, if $|\phi| \leq 1$ the AR(1) model will present itself as a mean reverting, stationary time series.

The implementation of such test, summarised by [12], is formulated as follows: The null hypothesis, $H_0 : \phi = 1$, states that the model has an unit root, meaning the series can be represented by a Random Walk process and, thus, the time series is non-stationary. The alternative hypothesis, $H_1 : |\phi| < 1$, states otherwise and therefore implies that the time series is stationary.

In order to perform the test, equation (2.18) is subtracted by y_{t-1} on both sides

$$\begin{aligned} y_t - y_{t-1} =& \phi y_{t-1} - y_{t-1} + \epsilon_t \\ y_t - y_{t-1} =& (\phi - 1) y_{t-1} + \epsilon_t \end{aligned} \tag{2.19}$$

yielding:

$$\Delta y_t = \beta y_{t-1} + \epsilon_t \tag{2.20}$$

where $\beta = (\phi - 1)$

The DF test statistic, t_{DF}, is calculated as follows

$$t_{DF} = \frac{\hat{\beta}}{S(\hat{\beta})} \tag{2.21}$$

where $\hat{\beta}$ is the least square estimate of β and $S(\hat{\beta})$ is the standard error of the estimate.

If the DF statistics is lower than a critical value, tabulated in [8], the null hypothesis, H_0 is rejected, therefore, the time series is stationary.

Later, in 1984, the test was extended by Said and Dickey [13] to ARMA (q,p) models with unknown order:

$$y_t = c + dt + \phi y_{t-1} + \sum_{j-1}^{p} \gamma_j y_{t-j} + \epsilon_t \tag{2.22}$$

where c is a constant and dt a trend term. This became known as the Augmented Dickey-Fuller test (ADF).

Similarly to the DF test, y_{t-1} is subtracted in both sides of (2.22), yielding:

$$\Delta y_t = c + dt + \beta y_{t-1} + \sum_{j-1}^{p} \gamma_j y_{t-j} + \epsilon_t \tag{2.23}$$

and the ADF statistics is computed as:

$$t_{ADF} = \frac{\hat{\beta}}{S(\hat{\beta})} \tag{2.24}$$

Once again, if the ADF statistics is lower than a critical value, the null hypothesis, H_0 is rejected and, therefore, the time series is stationary. If, instead, the ADF statistics is higher than the threshold, H_1 is rejected and the series is assumed to be a Random Walk which depicts an unpredictable process.

In our case, this test will be used to evaluate the stationarity of the spread series between the two assets which form each pair. Pairs deemed non-stationary are discarded and are not traded.

Upon the discussion of the concepts introduced in this chapter, we are in condition to begin examining the decision process behind the implementation of the strategy.

Chapter 3

Decision Making Process

Having presented and discussed, the essential theoretical concepts and methodologies, we now possess the necessary tools for developing a practical procedure for the implementation of a Statistical Arbitrage Pairs Trading strategy.

The approach adopted in this work can be divided into four main stages:

- selecting stocks,
- forming pairs,
- filtering pairs,
- applying an investment strategy to the filtered pairs.

All the steps of the list presented above, are to be executed resorting to historical data from stocks quoted in the US Stock Market. As a consequence, we must gather that data in a way that fits our purpose.

3.1 Data Processing

With that in mind, we will use Python software, and resorting to the Yahoo Finance library, we will automatically download the tickers (abbreviations used to uniquely identify publicly traded shares) from all S&P500 index's stocks. For each of these assets, daily information regarding their open, low, close, high and adjusted close prices is available. In addition, it is also possible to retrieve data concerning the volume, and eventual dividend yields, from each company across each day. However, for our case, we must properly filter this information, focusing our attention on each asset's price series.

In fact, historical adjusted close price data, regarding stocks from the S&P500 index, will be used to evaluate the entire algorithm. The S&P500 is a dynamic index, meaning that the stocks which form it are frequently changed. Assets are often added, or removed, based on the underlying company's valuation. Therefore, we will only consider stocks that were already quoted, in it, at the beginning of the strategy

,and which have remained quoted in it ever since (approximately 450 stocks). Securities with incomplete data are discarded.

In order to avoid data bias, the beginning of the strategy must be set in a particular point in time, and then, the first three points of the list presented above, will be executed assuming past price values in relation to the starting date. The fourth and last point of the list will require future data, in relation to the start date. As a consequence, through the course of this work, historical data must be divided in two groups: training data (required to execute the first three stages), and test data (required to apply and test the strategy). This principle is described on Figure 3.1.

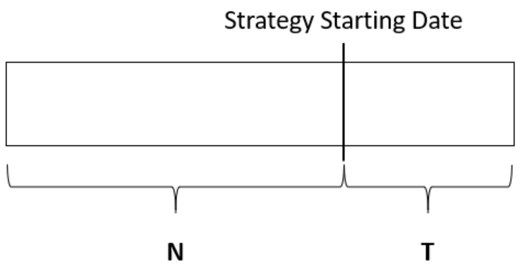


Figure 3.1: Price data division

where N represents the number of observations used for training the model, and T represents the number of observations used for testing it. If the starting date is set at day d, at that point in time, the investor only has access to data up until the end of day d.

A proper understanding of this approach is critical, since many important parameters and estimations, regarding the strategy's implementation, are time sensitive and depend on the time frame used for their calculation. Therefore, the number of observations considered for training, N, can considerably affect results.

On the other hand, conclusions taken upon training will be pursued during testing. Hence, the time frame considered for applying the strategy, T, is key to ensure the continuity of results. In other words, the deductions obtained in the training stage, have an "expiration date". From this expiration date onwards, previously acquired presumptions start to become obsolete. In addition, the strategy is partially passive, meaning some of its parameters, are set during training and do not change during testing.

As for starting dates, the system will be tested at 15 different points in time. The first test will begin on the first business day of 2017 (2017-01-03). From there, the 14 remaining tests will start 50 business days apart, into the future. The last test will begin on 2019-10-15. It is worth noticing that the US Stock Market is open around 252 days per year.

For each of the 15 starting dates, and for each of the selection methods, distinct combinations, of N, and T, will be considered and tested.

3.2 Selecting Stocks

In the starting stage of the implementation we first need to develop a method to properly select stocks which present a greater chance of success. Simply combining all S&P500 stocks, would result on hundreds of thousands of pairs. Computing cointegration tests on all these pairs is, in a real environment, very time consuming and not practical. As a consequence, three approaches for preselecting stocks are followed in this dissertation.

The first two (Section 3.2.1), take into consideration each stock's relation with the market (β_i), during the previous N days. Stocks with similar Beta values will be chosen and combined to form pairs. In order to do so, two groups of stocks will be formed. One containing the ten assets which present the higher values of Beta and other containing the ten assets whose Betas are closer to one. Stocks contained in each group will then be combined amongst each other to form pairs.

The third approach is related with the Hurst Exponent (Section 3.2.2). This statistical test allows for a comprehensive study of a time series' long term memory. In other words, the Hurst Exponent can be used to effectively quantify a given series' trend. Therefore, this parameter will be computed for each stock during the N day training period. The ten assets which present the lowest values (indicating lateral trends) will be selected and grouped. Once again, pairs will be formed by combining them amongst each other.

3.2.1 Selection based on stocks' relation with the market

The Capital Asset Pricing Model (CAPM)

The Capital Asset Pricing Model (CAPM) is one of the most widely accepted theoretical assumptions for modelling stock returns [14–20]. In order to understand it, first it is important to explore some related notions.

By computing a linear regression between the observed values of a given security's returns and the market's returns, over a certain period of time, we arrive at the following expression

$$r_i = \alpha_i + \beta_i r_m \tag{3.1}$$

where, α_i and β_i are the parameters of the linear regression, and r_i and r_m are the security and market returns in that period. From here, we can add terms on both sides of the equation and rewrite it in a way that will take the risk free rate into account. The result is as follows

$$r_i - r_f = \alpha_i - r_f + \beta_i r_f + \beta_i r_m - \beta_i r_f \tag{3.2}$$

The risk free rate, r_f, is the zero risk return rate at which an individual can invest money. For simplification reasons, this rate is often assumed to be equal to the interest rate at which an individual can borrow money. The reference value for r_f is usually around 3%.

Therefore, in the CAPM context, the assumption for the relation between the return of a given asset and the market is

$$r_i = r_f + \beta_i(r_m - r_f) \tag{3.3}$$

Equation (3.3) is known in the world of finance as Security Market Line (SML). It is easy to note that for this expression to be compatible with (3.1), in average $\alpha_i^* = \alpha_i + r_f(\beta_i - 1) = 0$. In this condition, in average, $\alpha_i = r_f(1 - \beta_i) \neq 0$. In practice, as we compute the linear regression, this might not be true. Only when we consider a large set of data, does this condition reveal itself as true for the average results.

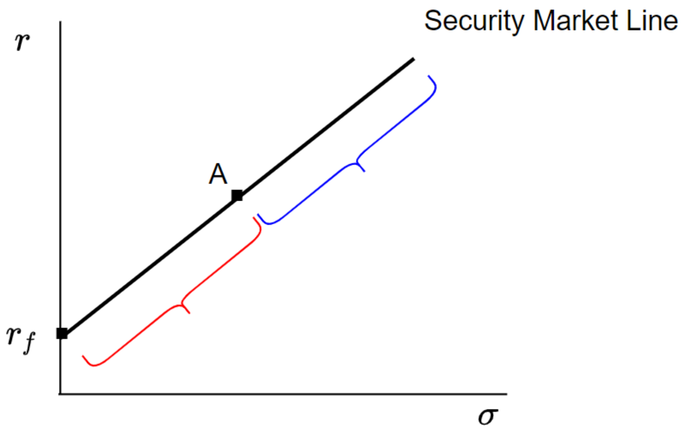


Figure 3.2: Security Market Line

Figure 3.2, where the x-axis of the chart represents the risk (in terms of standard deviation), and the y-axis of the chart represents the expected return on a given asset, depicts the Security Market Line. This line acts as a graphical representation of CAPM and can be useful to determine whether a particular financial product offers favourable expected returns when compared to its own level of risk. Point **A** represents an investment in a risky asset with a certain expected return, and a certain degree of risk. The part of the SML, highlighted in red, describes the evolution of the expected return, r, and the risk, σ, if a fraction of the capital was to be invested in a risk free product, and the rest was invested in **A**. The part of the SML, highlighted in blue, corresponds to borrowing money at the interest rate r_f and investing it in **A**.

However, for investment purposes, what should interest us is the relative variation of r_i and r_m which is given by

$$\frac{\partial r_i}{\partial r_m} = \beta_i \tag{3.4}$$

where β_i, widely known in the finance world as Beta, serves as a ratio indicator between market returns and a given asset's returns. For example, $\beta_i = 3$ is an indication that when the market is up by 1%, that security is likely to go up by 3%. Analogously, if a certain stock has $\beta_i = 0.5$, if the market moves by 10%, that stock is expected to move by 5% in the same direction as the market.

Selection Process

Quite obviously, the first step is to define a starting day for the strategy. Having established this, the training data must be set as the N past values of the price series until the starting date. For example, if we define the start of the strategy to be on the first day of 2018, the last N days of 2017 should be used for training. These two stages are to be repeated on all three approaches.

Once the training period has been properly specified, the computation of Betas may begin. As stated in section 3.2.1, β_i is a ratio factor between the return of a given asset and the return of the market, on the same time period. Considering that the S&P500 index represents the combined valuation of all stocks which form it, in our case, this index will be assumed to model the market behaviour.

This being said, we first need to compute the daily returns of both the market (S&P500 index) and each individual stock. This daily return, or daily percentage change, during day d, can be computed as

$$r_d = \frac{p_d - p_{d-1}}{p_{d-1}} \tag{3.5}$$

where r_d is the daily return on that security, p_d is the security's adjusted close price on day d, and p_{d-1} is the security's adjusted close price on day $d-1$.

Once the transformation from daily prices to daily returns is complete, a linear regression is computed between the returns on the market, and the returns on the stock. The resulting regression is expressed as

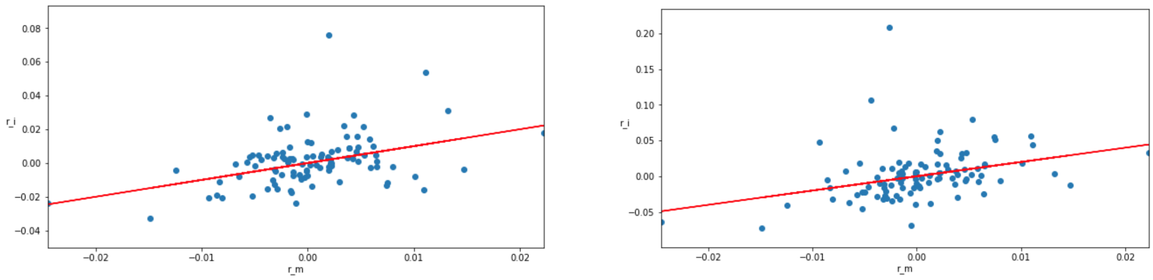
$$r_i = \beta_i r_m + \alpha \tag{3.6}$$

where α and Beta, (β_i), are the regression's parameters. α is usually very small and is assumed to be zero. More importantly, Beta, β_i, can be computed as follows

$$\beta_i = \frac{COV(r_m, r_i)}{VAR(r_m)} \tag{3.7}$$

where r_m and r_i are, respectively, the market's and security's, daily return series.

Figure 3.3 (a), presents the linear regression between the Wells Fargo Bank stock's (WFC) daily returns, and the S&P500 Index's daily returns, during the last 100 business days of 2016. On its right, Figure 3.3 (b), displays the same linear regression, performed during the same time period and using the same market model, but resorting to data referring to Marathon Oil Corporation (MRO).



(a) Linear Regression between WFC stock's daily returns and S&P500 Index's daily returns, N=100.

(b) Linear Regression between MRO stock's daily returns and S&P500 Index's daily returns, N=100.

Figure 3.3: Linear regression between WFC stock's daily returns (left), MRO stock's daily returns (right), and market daily returns, using data from 2016-08-10 to 2016-12-30 ($N = 100$).

In these two cases, the asset's Betas, which can be interpreted as the slope of the regression model presented in red, have a value of $\beta_{WFC} = 1.00$, for the Wells Fargo Bank stock, and $\beta_{MRO} = 2.00$, for the Marathon Oil Corporation stock.

This process is repeated for all stocks quoted in the S&P500 Index, and all corresponding Betas are computed. The Betas are then sorted in descending order. The 10 stocks with the highest values are chosen, as well as the ones whose values are closer to one, and two groups are formed. Assets from each group will be combined amongst each other originating 45 pairs per group. The method for creating these pairs will be later discussed in section 3.3.

3.2.2 Selection based on stocks' Hurst Exponent value

The Hurst Exponent (H)

The Hurst exponent (H) is a statistical test used to measure the long-term memory of a time series. Since its introduction by Hurst [21] in 1951, it has been widely used in the field of finance [22–24]. In 2017, Ramos-Requena et al. [25], successfully applied the concept to Pairs Trading. The authors select pairs by computing their respective spread series' Hurst Exponent value. Pairs with the lower spread series' Hurst, are then selected for trading. In our case, the Hurst Exponent, of each individual stock, shall be considered as a selection criterion.

Based on the value of H, a time series can be classified as (1) anti-persistent ($0 < H < 0,5$), (2) uncorrelated ($H = 0,5$) or (3) persistent ($0,5 < H < 1$). In other words, a persistent series is one where a clear trend, either bullish or bearish, can be observed. As for the anti-persistent case, series where $0 < H < 0,5$ are usually representative of markets with an horizontal trend (neither bullish nor bearish). Lastly, uncorrelated time series with $H = 0,5$ indicate a Random Walk process which, by definition, is unpredictable.

H can be estimated based on a Rescaled Range analysis, also known as R/S analysis, and it is defined as

$$\frac{R}{S}_n = cn^H \qquad \text{as } n \to \infty \tag{3.8}$$

where $\frac{R}{S}_n$ is the Rescaled Range, c a constant, n the number of observations and H the Hurst exponent.

The first step in the estimation of H consists in the said Rescale Range analysis, which is addressed in this dissertation using the same notation followed by Couillard and Davison in [26].

First, the time series of length N is divided into M "sub-time series" of length n such that $M \times n = N$. Each "sub-time series", or chunk, is designated as I_m, with $m = 1, 2, ..., M$ and their corresponding data points are labelled as $N_{k,m}$, where $k = 1, 2, ..., n$.

Subsequently, for each chunk, I_m:

1. Calculate mean value μ_m

$$\mu_m = \frac{1}{n} \sum_{k=1}^{n} N_{k,m} \tag{3.9}$$

2. Calculate standard deviation S_{I_m}

$$S_{I_m} = \sqrt{\frac{1}{n} \sum_{k=1}^{n} (N_{k,m} - \mu_m)^2} \quad \text{for } k = 1, 2, ..., n \tag{3.10}$$

3. Create the mean-centred series

$$X_{k,m} = N_{k,m} - \mu_m \qquad \text{for } k = 1, 2, ..., n \tag{3.11}$$

4. Calculate the cumulative sum of mean-centred series

$$Y_{1,m} = X_{1,m} \tag{3.12}$$

$$Y_{k,m} = Y_{k-1,m} + X_{k,m} \qquad \text{for } k = 2, ..., n \tag{3.13}$$

5. Create the range series R

$$R_{I_m} = max(Y_{k,m}) - min(Y_{k,m}) \quad \text{for } k = 1, 2, ..., n \tag{3.14}$$

6. Calculate the Rescale Range for a given value of length n

$$\frac{R}{S}_n = \frac{1}{M} \sum_{m=1}^{M} \frac{R_{I_m}}{S_{I_m}} \tag{3.15}$$

The previous steps are illustrated by Fig. 3.4, for a time series of length $N = 10$ and by choosing $n = 5$. However, it is important to clarify that this analysis expects the use of several values of n.

Finally, the Hurst exponent (H) can be estimated using a linear regression by applying the log in both members of (3.8), yielding:

$$log\left(\frac{R}{S}_n\right) = log(c) + Hlog(n) \tag{3.16}$$

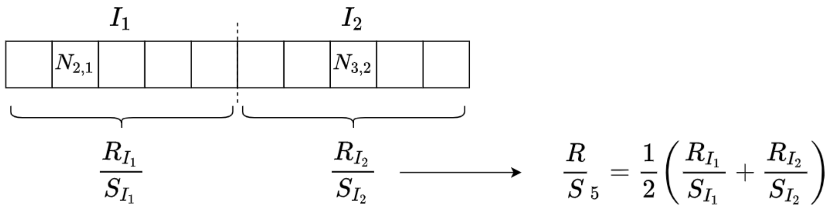


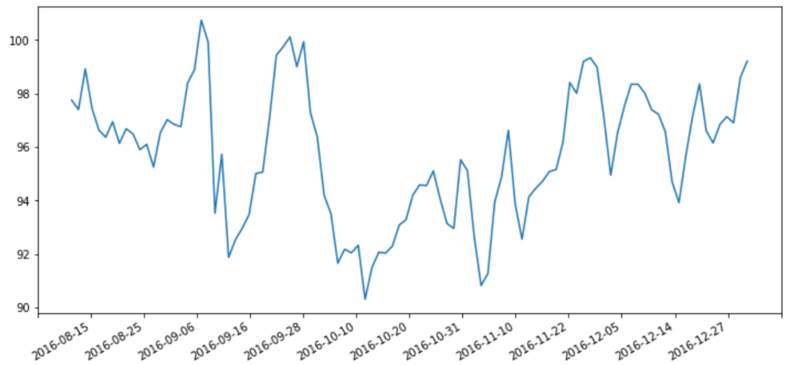
Figure 3.4: Example of the rescale range analysis of a time series of length N=10, using n=5

Selection Process

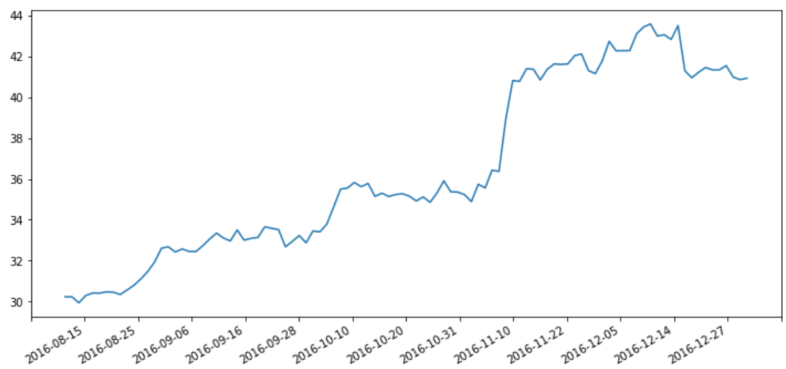
Once again, we first require the setting of a start date. Furthermore, tests performed using this selection method, will begin on the exact same 15 dates as the ones discussed above.

In this section, anti-persistent time series were defined as series whose Hurst Exponent value was between 0 and 0.5 ($0 < H < 0.5$). In these cases, the time series are expected to depict a "trendless" behaviour, or lateral trend. That is, during the considered studying period, the series did not diverge from a given mean value, and neither revealed a bullish nor a bearish trend.

On Figure 3.5, are presented two stock price series from 2016-08-10 to 2016-12-30. The first image (a), illustrated above, displays the price of Alexandria Real Estate Equities Inc's stock (ARE) during this time frame. The second image (b), represented below, portraits the price of Metlife Inc's stock (MET) in the same period.



(a) Alexandria Real Estate Equities Inc's stock price from 2016-08-10 to 2016-12-30, N=100.



(b) Metlife Inc's stock price from 2016-08-10 to 2016-12-30, N=100.

Figure 3.5: Alexandria Real Estate Equities Inc's (a) and Metlife Inc's (b) stock prices from 2016-08-10 to 2016-12-30 ($N = 100$).

As we can see, both series depict very different behaviours. On Figure 3.5 (a), the series does not reveal any clear trend and it appears to reveal a mean reverting process. Contrarily, on Figure 3.5 (b), the series reveals a clear bullish trend and the prices appear to be diverging from the mean. The Hurst Exponent value of each asset throughout this time period was $H_{ARE} = 0.30$ for the ARE stock and $H_{MET} = 0.92$ for the MET stock.

As for our case, in order to implement a statistical arbitrage pairs trading strategy, it may be useful to select financial products which do not diverge and do not denote any clear trend. As a consequence, only stocks with Hurst Exponent values between 0 and 0.5 will be considered.

In the course of this work, the Hurst Exponent will be calculated, at each starting date, and using different N time intervals, for all stocks quoted in the S&P500 Index. Once they are computed, stocks will be ordered according to their value of H. The ten assets with lower Hurst Exponent values will then be selected and grouped. As for previous selection methods, these ten stocks will be combined amongst each other forming 45 different pairs.

3.3 Forming Pairs

In the previous stage, we have chosen ten stocks which, a priori, present greater chances of being combined amongst each other to form pairs whose spread series is stationary. In fact, for each trading period's starting date, and for each N day interval used for training, three groups of ten stocks were formed. One containing the ones selected based on high correlation values with the market, another containing stocks with unit values for Beta, and a third containing the stocks whose Hurst Exponent values are the lowest. For each group of selected stocks, all possible forty five pairs ($\binom{10}{2} = 45$) will be formed. The order of the stocks will not be taken into account.

3.3.1 Properties of the Spread

Having established the assets to be paired, the computation of their respective spread series may begin. The spread series between a given pair of financial products X/Y, V_P, can be defined as

$$V_P = P_y - BP_x \tag{3.17}$$

where P_y is the price series of asset Y, P_x is the price series of asset X and B is a weight constant between the two stocks.

As previously explained, any statistical arbitrage pairs trading strategy, involves the simultaneous opening of both a short, and a long position, in a given pair of financial products. By doing so, the spread between the prices of both stocks becomes the investment product. As a consequence, two positions can be adopted by the investor: to either long or short the spread. In our case, in order to long the spread, the investor must buy stocks from company Y and short sell stocks from company X (betting for the increase of V_P). As for shortening the spread, the trader must do the exact opposite: Short sell Y's stocks and buy X's stocks (betting for the decrease of V_P). Furthermore, the weight coefficient, B, will determine the relative amount of investment in each asset. For example, if the investor wishes to long the spread, for each Y stock bought, he must short sell B stocks from X.

As shown in equation (3.17), the construction of the pair's spread is performed resorting to price data from both stocks. By computing the linear regression between both price series, and assuming stock X as the independent variable, the expected value for the price of stock Y, $\hat{P}_y$, can be expressed as

$$\hat{P}_y = A + B.P_x \tag{3.18}$$

where A and B are the regression's parameters. The price of security Y, P_y can then be defined as

$$P_y = \hat{P}_y + P_{y_e} \tag{3.19}$$

where P_{y_e} is an error term. In the absence of deviations, equations (3.19) and (3.17) can be rewritten, respectively, as

$$P_y = \hat{P}_y \tag{3.20}$$

$$\Delta VP = \Delta P_y - B\Delta P_x = 0 \tag{3.21}$$

As for the spread, its expected value, $\hat{V}_P$, can be computed as

$$\hat{V}_P = \hat{P}_y - BP_x = A \tag{3.22}$$

Similarly as for the price series, the error term for the estimation of the spread, V_{P_e}, can be calculated as

$$\begin{aligned} V_{P_e} &= V_p - \hat{V}_P \\ &= V_P - A \\ &= P_{y_e} \end{aligned} \tag{3.23}$$

which is the same as the error term of the price series.

Another very important property of the spread series has to do with its variance. The mean value of the spread series, V_{P_m}, can be expressed as

$$V_{P_m} = P_{y_m} - BP_{x_m} \tag{3.24}$$

where P_{y_m} and P_{x_m} are the mean values of the Y's stock price series and the X's stock price series, respectively. The mean value of the spread series, will be greater than zero if the average price of security Y is higher than the average price of B stocks from asset X, and smaller than zero if the opposite occurs.

The Variance of the spread series can now be defined as

$$VAR(V_P) = \frac{1}{N} \sum (V_P - V_{P_m})^2 \quad (3.25)$$

where N is the number of observations in each price series.

Taking into account equation (3.17), the variance of the spread can be computed, using data from both stock's price series, as

$$Var(V_p) = Var(P_y) + B^2.Var(P_x) - 2BCov(P_x, P_y) \quad (3.26)$$

It is easy to note that, the lower the covariance between both price series is, the higher the variance of the spread series will be. Therefore, decreasing covariance between two assets, leads to an increase on the spread's variance and, thus, can positively impact the pair's profit prospectives.

3.3.2 Spread computation

In order to provide a clear and easy to follow methodology for spread computation, an example will be discussed. Figure 3.6 presents the quotation of two stocks from the S&P500 index. The stock price of company Advanced Micro Devices (AMD) is presented in blue, and the stock price of United Rentals, Inc. (URI) is presented in red. The value of AMD's stock is shown on the left y axis, and the URI's stock price is shown on the right y-axis. Both series refer to a period between 2016-01-06 and 2016-12-30.

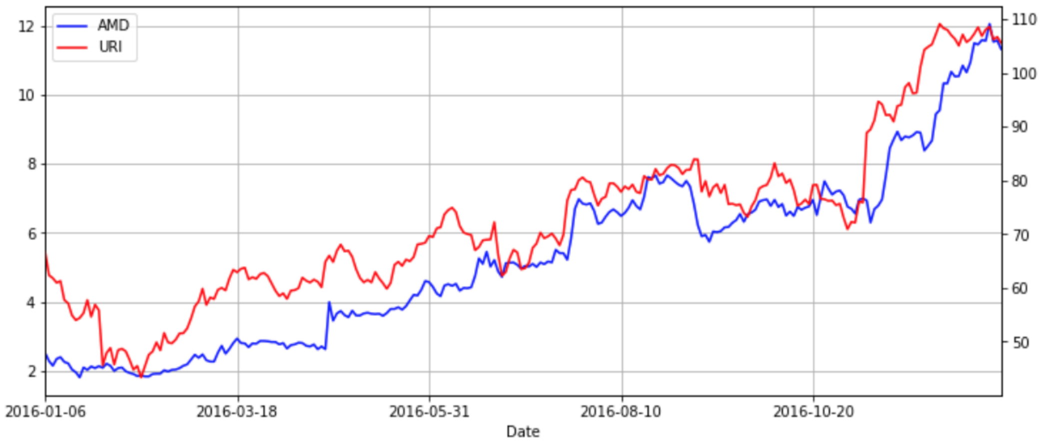


Figure 3.6: Prices of Advanced Micro Devices' stock (AMD), and United Rentals, Inc.'s stock (URI), during the last 250 business days of 2016 ($N = 250$)

Both these stocks were chosen, upon the selection stage, using the high Beta criterion. During the 250 day period, AMD revealed a Beta of $\beta_{AMD} = 2.10$ while URI revealed a Beta of $\beta_{URI} = 2.04$.

In pursuance of constructing the spread, a linear regression between the prices of the two assets, throughout the training period (from 2016-01-06 to 2016-12-30), can be computed. Figure 3.7, presented below, depicts this linear regression.

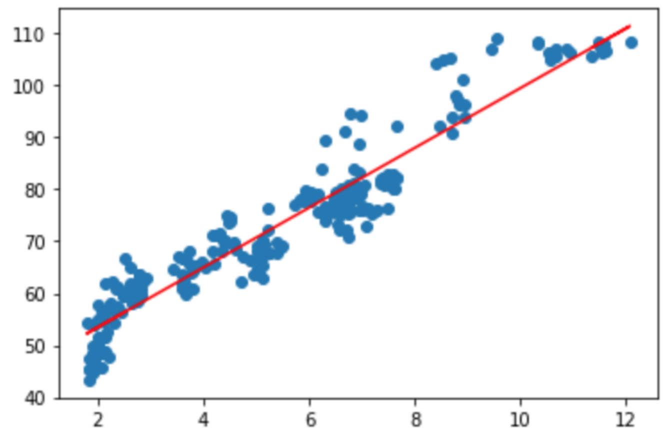


Figure 3.7: Linear Regression between the prices of Advanced Micro Devices' stock (AMD) and United Rentals, Inc.'s stock (URI), during the last 250 business days of 2016 ($N = 250$)

The resulting weight coefficient, B, can be calculated as

$$B = \frac{COV(P_{(AMD)}, P_{(URI)})}{VAR(P_{(AMD)})} \tag{3.27}$$

where $P_{(AMD)}$ and $P_{(URI)}$ denote the price series for both stocks. In this particular case, the value of parameter B was computed as $B_{Pair} = 5.76$. In practice, this value can be interpreted as a valuation ratio between the two assets. In other words, in average, each time the AMD stock values by one dollar, the URI stock is expected to value by B dollars.

Having set B, it is now possible to form the spread series of the pair. Figure 3.8 presents the resulting spread series of the pair AMD/URI from 2016-01-06 to 2016-12-30.

Figure 3.8: Spread series between the prices of Advanced Micro Devices' stock (AMD), and United Rentals, Inc.'s stock (URI), during the last 250 business days of 2016 ($N = 250$), with $B = 5.76$

As it is easily noticeable, the spread series of this pair appears to reveal a stationary, mean reverting behaviour. However, performing a detailed visual graphic analysis on all spread series is an impossible mission. As a consequence, a proper filtering methodology must be implemented in order to select pairs whose spread series disclose such characteristics.

3.4 Filtering Pairs

As previously referred in Section 2.3, the Augmented Dickey-Fuller test (ADF), can be a valuable tool in determining if a given series is stationary.

The ADF test assumes an initial null hypothesis regarding the existence of an unit root in the time series. If this null hypothesis is rejected, than the series is said to have a stationary, mean reverting behaviour. If, instead, the null hypothesis is confirmed, the time series is assumed to behave as an unpredictable, non-stationary, Random Walk model.

In order to test this hypothesis, a parameter called ADF statistics is computed. The calculated value of this statistic will determine the result of the test.

All pairs (45 for each starting date, for each selection method, and for each N day training period), will undergo an ADF test. After being computed, the resulting ADF statistics of each pair's spread series, are compared with the critical values presented on the table below.

Probability (%)	90	95	99
Critical value	-2,57	-2,87	-3,44

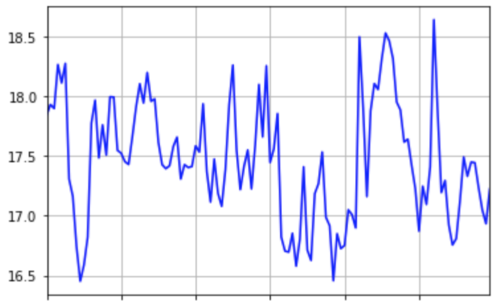
Table 3.1: Critical values for the ADF test statistics.

These critical values mark the boundaries of the confidence intervals for the test's results. If the ADF statistics is equal to -2.57, it is possible to state that there is a 90% probability of correctly rejecting the null hypothesis. For the purpose of this work, the threshold for the ADF statistics will be set as $thr = -3.44$, based on the 99% confidence value. Once tests are ran for all pairs, the resulting statistics are compared with the threshold value.

Pairs whose statistics are equal or lower than the predetermined threshold, are assumed to have stationary spread series and, consequently, are led into a trading period. Pairs whose statistics are higher than the predetermined threshold are excluded and discarded.

Figure 3.9 presents two plots of two different spread series, concerning two distinct asset pairs. Both spreads were computed by pairing stocks which were selected using the high Beta method. Furthermore, the price data used to construct the model was collected during the same period (last 120 business days of 2016), for all four stocks. On the left image (a), the spread series between NRG Energy Inc's ($\beta_{(NRG)} = 2.31$) stock price and Freeport-McMoRan's ($\beta_{(FCX)} = 2.48$) stock price is shown. On the right image (b), it is possible to observe the spread series between United Rentals, Inc's ($\beta_{(URI)} = 2.31$) stock price and Franklin Templeton Investments' ($\beta_{(BEN)} = 2.48$) stock price.

(a) NRG/FCX spread series between 2016-07-13 and 2016-12-30 ($N = 120$), $B_{Pair} = 0.12$.



(b) URI/BEN spread series between 2016-07-13 and 2016-12-30 ($N = 120$), $B_{Pair} = 0.14$.

Figure 3.9: NRG/FCX's (a) and URI/BEN's (b) spread series between 2016-07-13 and 2016-12-30 ($N = 120$).

As we can see, the spread series of the NRG/FCX pair appears to be diverging from its mean value, while the spread series from the URI/BEN pair appears to be constantly reverting towards it. It is interesting to note that the ADF statistics results for each pair were: $t_{(NRG/FCX)} = -1.44 > -3.44$ and $t_{(URI/BEN)} = -4.75 < -3.44$. As a consequence, the pair NRG/FCX was discarded while the pair URI/BEN successfully passed the filtering stage and was admitted in the testing phase.

3.5 Applying an investment strategy to the filtered pairs

Now that we have determined the asset pairs which depict a long lasting price equilibrium between them, and whose resulting spread series revealed a stationary, mean reverting behaviour, we can use them in applying a Statistical Arbitrage Pairs Trading strategy.

This is the fourth and last stage of the implementation. As explained on Figure 3.1, all the trading periods in this section, were simulated using future data in relation to the training period. In reality, at the end of trading day t, the investor only has access to data up until that moment.

At this point, it is important to reinforce the idea that this is a purely academic work. A segmented, step by step approach, will be developed, explained and analysed. However, several important simplifications and assumptions were adopted in relation to a real life investment environment. Decisions are made and executed while the stock market is closed (during the night), and positions are immediately opened in the morning of the following day. The entry price for these positions is assumed to be equal to the previous day's adjusted close price. In addition, all transaction fees (both for entering and exiting the market) are neglected. The bid-ask spread is also assumed to be zero. Capital costs are not considered either. As a consequence, this dissertation does not aim to provide any guarantee paths for profit but, instead, attempts to formalize practical guidelines for the implementation of statistical arbitrage pairs trading strategies.

As previously explained (Section 2.1), such strategies involve the simultaneous opening of both a long and a short position in each of the securities which form the pair. In other words, the investment product of the strategy is the spread series of the pair, and the investor can choose to either short it or long it. The opening of these positions requires a substantial deviation of the series from its mean value. When this happens, and assuming that the mean reverting behaviour demonstrated in the past will continue, the investor must enter a position which will allow him to profit if mean reversion, in fact, occurs.

Figure 3.10, presented below, illustrates the basic methodology behind any statistical arbitrage pairs trading strategy. On it, drawn in blue, is a generic spread series, from a given pair of financial products (X/Y). The image depicts a 200 day trading period for this pair. Three horizontal lines, one in black and two in red, are also portraited on the figure. The first, represents the mean value of the spread series. The two latter, represent the threshold values for the opening of positions, and are based on the series' standard deviation. Every time the spread deviates from the mean, by more than this threshold value, a signal is generated and a position is opened. Once the spread reverts to its mean, the position is withdrawn. The opening and closing of positions are also identified on the image by green and red dots, respectively.

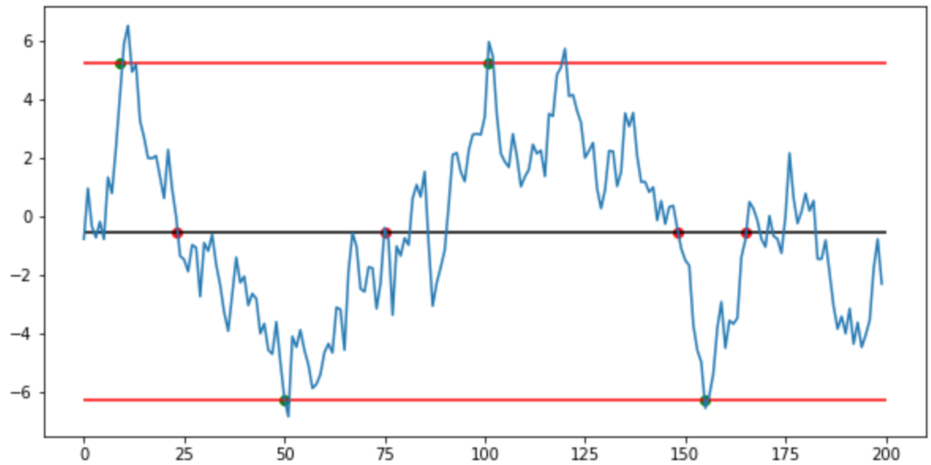


Figure 3.10: Generic Pairs trading strategy example

As we can see, the first green point (9^{th} day), marks the first opening of a position. In this case, the spread deviated, by excess, from the mean. Therefore, a signal for shortening the spread was generated. In other words, and according to the definition of the spread (Equation (3.17)), the investor must long B stocks from asset X, for each stock from asset Y he shorts. Once the spread shortened and reverted to its mean, the first position was withdrawn (red dot on the 23^{rd} day). From this point onwards, three other positions were opened (the three remaining green points on the 50^{th}, 101^{st} and 155^{th} day). However, in the cases of the positions opened on the 50^{th} and 155^{th} days, the spread deviated by default from the mean. As a consequence, a signal for longing the spread is generated. Inversely to shortening the spread, in these cases, the investor must short B X stocks for each Y stock he buys. As before, as soon as the spread reverts to its mean value, the positions are withdrawn and profit is taken (red points on the 75^{th} and 165^{th} day).

3.5.1 Mean reversion and Z-Score

Despite being helpful to understand pairs trading, Figure 3.10 is not an accurate representation of a real strategy. In reality, we can not just simply compute the spread's mean and standard deviation during the trading period. Since they are based on future data, unknown to the investor at the moment of decision, another methodology must be adopted. In our case, the series' mean and standard deviation, are to be set as a moving average mean of its 15 previous values. Hence, the reference value for the spread series' mean, at day t, μ_t, can be set as

$$\mu_t = \frac{\sum_{k=0}^{14} V_{P(t-k)}}{15} \tag{3.28}$$

where $V_{P(t-k)}$ is the value of the spread at day $t-k$. Similarly, during the trading period, the reference value for standard deviation of the series, at day t, σ_t, is computed as

$$\sigma_t = \overline{\sigma}_{daily}[t-14;t] \tag{3.29}$$

where $\overline{\sigma}_{daily}[t-14;t]$ is the average daily standard deviation of the spread series during the previous 15 days.

Figure 3.11, presented below, serves as a visual support for this implementation. On it, a 45 day trading period between 2017-03-16 and 2017-05-19 is shown. The line in blue represents the spread series between United Rentals, Inc (URI) and Advanced Micro Devices (AMD) with $B = 0.11$. The line in red represents the 15 day moving average mean of the spread series. The two green lines serve as the threshold benchmarks for the opening of positions. The upper green line is defined, at each point in time, as $\mu_t + \alpha \times \sigma_t$ while the lower green line is computed as $\mu_t - \alpha \times \sigma_t$, where α is a constant.

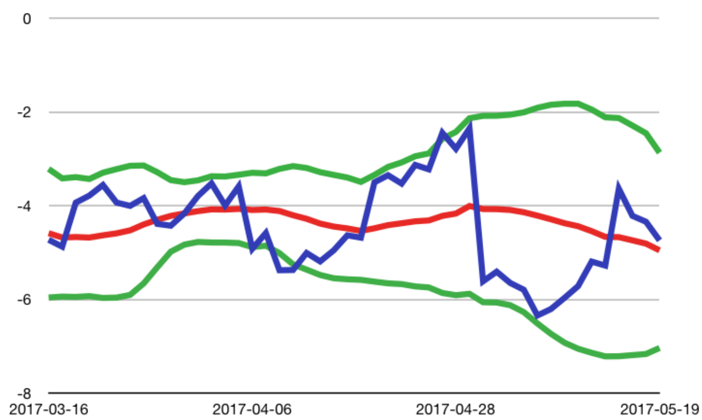


Figure 3.11: URI/AMD spread series between 2017-03-16 and 2017-05-19 (blue), respective 15 day moving average mean (red) and upper and lower threshold bands (green).

As in the previous, more simplistic approach, represented on Figure 3.10, every time the spread crosses the lower or upper bands of the threshold, a position is opened. Once the spread reverts to its mean, the positions are withdrawn and profit is taken.

Nonetheless, for computational reasons, it is important to introduce the concept of Z-Score. The Z-

Score, is a statistical concept often used for normalizing distances. In our case, it serves as a measure of the spread series' deviation from its own mean. In fact, and more precisely, it quantifies the number of standard deviations the spread has deviated from the mean.

The Z-Score can be calculated as

$$z_t = \frac{V_{P(t)} - \mu_t}{\sigma_t} \tag{3.30}$$

where z_t is the value of the Z-Score, $V_{P(t)}$ is the value of the spread, μ_t is the value of its moving average mean and σ_t is the moving average standard deviation, at the end of trading day t.

Figure 3.12, presented below, depicts the Z-Score plot for the URI/AMD pair's trading period of the previous example (between 2017-03-16 and 2017-05-19).

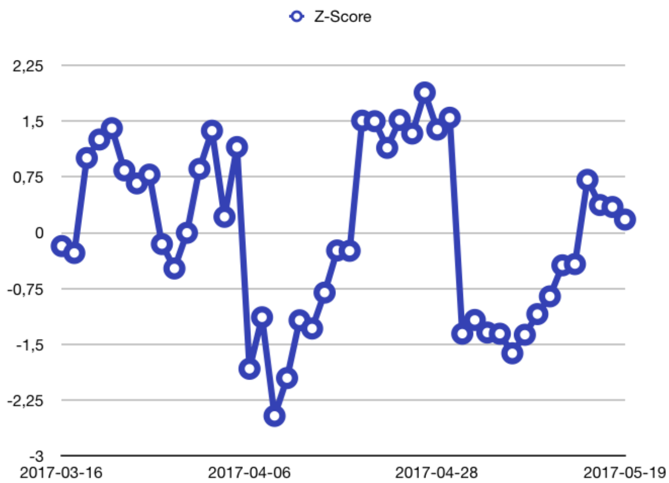


Figure 3.12: URI/AMD spread series' Z-Score between 2017-03-16 and 2017-05-19.

As we can see, the Z-Score allows for a normalization of the spread's deviation from the mean. As a consequence, a positive Z-Score indicates that the spread is above the moving average mean, while a negative Z-Score indicates that it is below it. If $z_t = 0$, then the spread at trading day t, is equal to the 15 day moving average mean. In fact, a Z-Score of $z_t = 1.5$, indicates that the spread is above the mean by 1.5 standard deviations ($V_{P(t)} = \mu_t + 1.5 \times \sigma_t$). On the other hand, a negative Z-Score denotes a deviation by default. For instance, a Z-Score of $z_t = -2$, indicates that the spread is below the mean by 2 standard deviations ($V_{P(t)} = \mu_t - 2 \times \sigma_t$).

Having established the concept of Z-Score, it is now possible to set the upper and lower thresholds for the opening of positions, as a function of the standard deviation.

3.5.2 Opening and closing positions

The upper and lower bands for the Z-Score can be defined as $z_t = \alpha$ and $z_t = -\alpha$, respectively. α is a constant and serves as a measure of how many standard deviations the spread must deviate from the mean, before a position is opened. For the purpose of this work, α was set equal to 1.75.

Figure 3.13, presented below, contains the exact same Z-Score from the previous example, as well as two lines (shown in red) which represent the upper and lower threshold values for entering the market, for $\alpha = 1.75$.

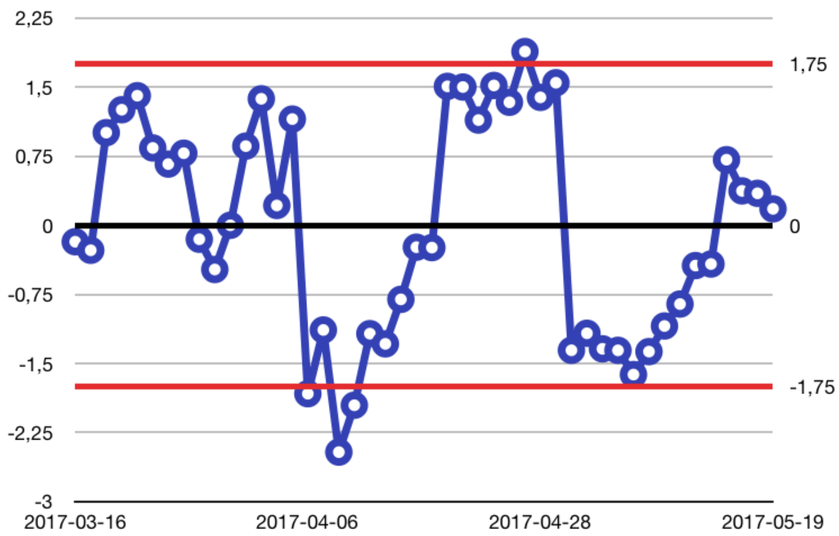


Figure 3.13: URI/AMD spread series' Z-Score between 2017-03-16 and 2017-05-19 (blue) and respective upper and lower threshold bands (red).

In order to control and execute, the opening and closing of positions, two types of signals must be generated:

- Long signals,
- Short signals.

Long signals are responsible for managing the opening and closing of long positions on the spread series, while short signals perform the same function when shortening the spread. The decision to open or close a position (for both longing and shortening the spread) relies on the value of these signs.

The long signal, at trading day t, $L_S(t)$, can assume three different values:

$$L_S(t) = 1 \vee L_S(t) = 0 \vee L_S(t) = none \tag{3.31}$$

If the Z-Score at trading day t (z_t), is below the average by more than 1.75 standard deviations ($z_t < -1.75$), the long signal is set equal to 1. If, instead, the Z-Score is positive ($z_t > 0$), the corresponding long signal is set equal to zero. Lastly, if none of the previous conditions applies (meaning that $-1.75 < z_t < 0$), the long signal adopts the $none$ setting.

Having established the proper generation of long signals, we can now execute the opening and closing of long positions. Long positions can take two forms:

$$L_P(t) = 1 \vee L_P(t) = 0 \tag{3.32}$$

where $L_P(t) = 1$ indicates an open long position at trading day t, while $L_P(t) = 0$ indicates that there are no active long positions, and that the investor is not exposed to the market. The value of the long position ($L_P(t)$) depends on the value of the long signal ($L_S(t)$). If $L_S(t) = 1$ or $L_S(t) = 0$, the long position assumes the same value as the long signal ($L_P(t) = L_S(t)$). If $L_S(t) = none$, the long position remains equal to its previous value ($L_P(t) = L_P(t-1)$). As a consequence, every time the Z-Score is lower than -1.75, and no active position exists, a long position is opened. This position remains open, and it is only closed when the Z-Score crosses its mean, and becomes positive ($z_t > 0$). Long signals, and corresponding long position for the example on Figure 3.13 are shown below.

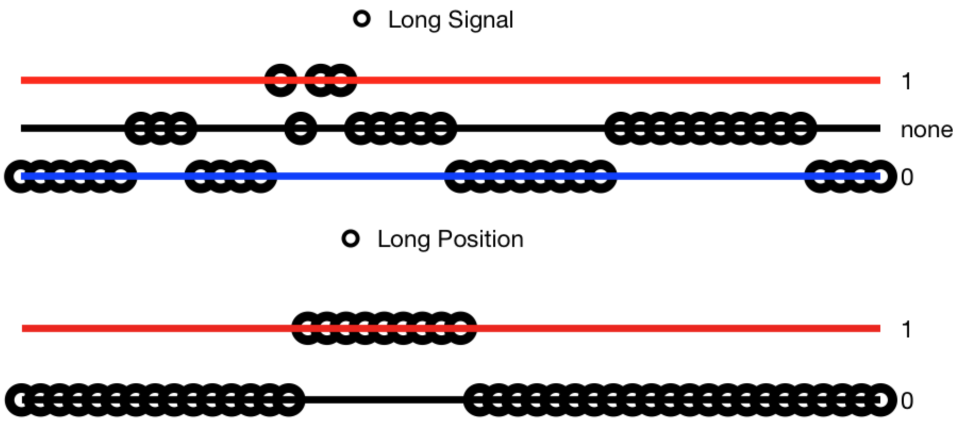


Figure 3.14: URI/AMD testing session's long signals and long positions, at the end of each day, between 2017-03-16 and 2017-05-19

Analogously, short signals can be set in three different ways:

$$S_S(t) = -1 \vee S_S(t) = 0 \vee S_S(t) = none \tag{3.33}$$

The procedure for setting the value of the short signal at day t ($S_S(t)$) is identical to the one used for the long case. However, if the Z-Score, at the end of day t is higher than 1.75, the short signal will be set as $S_S(t) = -1$. If the Z-score, at that moment, depicts a negative value, $S_S(t) = 0$, otherwise $S_S(t) = none$ is assumed.

$$S_P(t) = -1 \vee S_P(t) = 0 \tag{3.34}$$

Equation (3.34), presented above, describes the possible values for the short position at the end of day t ($S_P(t)$). Figure 3.15, displays the values of the short signal and short position, across the previous example's trading period.

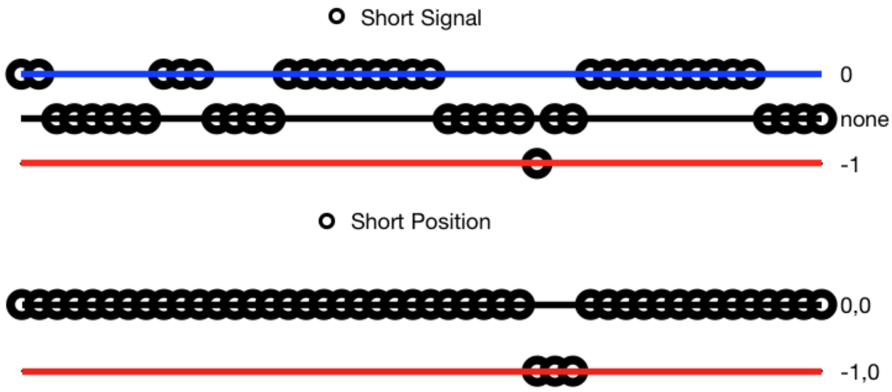


Figure 3.15: URI/AMD testing session's short signals and short positions, at the end of each day, between 2017-03-16 and 2017-05-19

Once again, the methodoly for its setting is similar to the previous case. If $S_S(t) = -1 \vee S_S(t) = 0$, then $S_P(t) = S_S(t)$. If, instead, $S_S(t) = none$, than $S_P(t) = S_P(t-1)$. This way, every time the spread is above the mean by more 1.75 standard deviations, a short position is opened, and it remains so, until the spread crosses the mean again.

Having set the short and long positions, it is now easy to define the strategy's position status at the end of day t, $P(t)$, as

$$P(t) = S_P(t) + L_P(t) \tag{3.35}$$

At this point, it is important to notice that there is an alternative way for closing both long and short positions. If, at the end of the T day trading period, an active position exists ($P(T-1) = 1 \vee P(T-1) = -1$), it will be automatically closed ($P(T) = 0$), independently from the Z-Score's value.

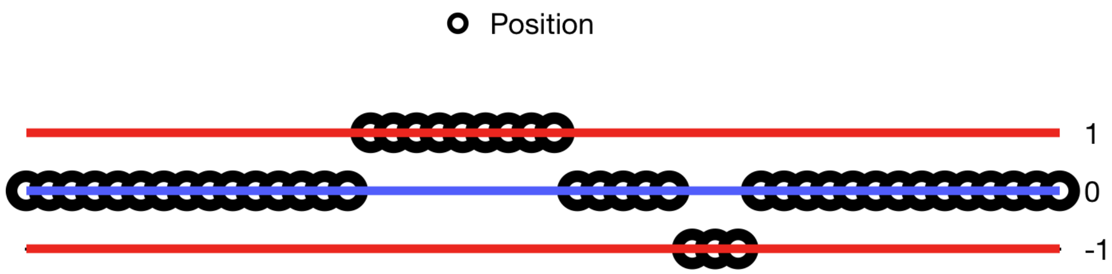


Figure 3.16: URI/AMD testing session's position, at the end of each day, between 2017-03-16 and 2017-05-19

As we can see, for the example represented on Figure 3.13, two positions were opened during the 45 day trading period. The first, was a long position, opened at the 16^{th} investment day (the first time the Z-Score was lower than -1.75), and closed at the 25^{th} investment day (when the Z-Score turned positive). The second, was a short position, opened at the 30^{th} trading day ($z_{30} > 1.75$), and closed at the the 33^{rd} ($z_{33} < 0$).

Having defined the entire decision process, it is now time to focus our attention on monitoring the strategy, and quantifying its performance. With that in mind, in the following chapter we will introduce several indispensable performance monitoring mechanisms.

Chapter 4

Performance Monitoring

At this stage, it becomes imperative to clarify some practical details regarding this strategy's implementation. Throughout the previous chapters, we have often referred to buying and short selling stocks. In reality, however, this model will be based on the trading of Contracts for differences (CFDs). A CFD is a contract between an investor and a broker, which allows the investor, to bet as to whether the underlying asset's price will rise, or fall. Differences between the open and closing trades are cash settled, without the transaction of any physical goods. CFDs offer investors all the benefits of owning a security, without actually having to own it.

More importantly to our case, this type of financial product, can often be margin traded, meaning the broker allows investors to borrow money to increase leverage. As brokers only require a deposit, to cover for possible losses, the minimum margin for opening such an account is usually between 15% and 20% of the total investment. For the purpose of this dissertation, a margin of 20% will be assumed.

4.1 Margin Trading Monitoring

The theoretical total investment (V_T) required for the opening of a long short position on a given pair of assets (X/Y) can be computed as

$$V_T = P_y + BP_x \tag{4.1}$$

where P_y is the buy price of security Y, P_x is the sell price of security X, and B is the weight coefficient of the pair. However, since the broker only demands 20% margin of this value for opening each position, the initial investment (V_0) can be set as

$$V_0 = V_T \times m_r \tag{4.2}$$

where $m_r = 0.2$ is the typical margin requirement of the broker. This capital (V_0), is expected to withstand eventual losses, from trades executed in that account.

In this work, each traded pair will be assumed to represent a single account, with a V_i initial investment. If, throughout the trading period, the margin is lost, and the account's balance goes to zero, the investor is liquidated and the pair does not trade any further. As a consequence, each pair's margin account's balance must be constantly evaluated.

In order to properly monitor the performance of the margin accounts, the first thing we need to do, is to define the daily spread differences during the trading period. The daily spread change during trading day t ($\Delta V_{P(t)}$), can be computed as

$$\Delta V_{P(t)} = V_{P(t)} - V_{P(t-1)} \tag{4.3}$$

where $\Delta V_{P(t)}$ represents the spread's change between the end of trading day $t-1$ and the end of trading day t. Quite obviously, this difference, only becomes effective when an active position was set at the end of trading day $t-1$. Therefore, the pair's account's balance change during trading day t ($DV(t)$) can be calculated as

$$DV(t) = \Delta V_{P(t)} \times P(t-1) \tag{4.4}$$

where $P(t-1)$ is the position in effect, on the previous day.

Hence, we can compute the account's balance at the end of trading day t ($V(t)$) as

$$V(t) = V_i + \sum_{q=0}^{t} DV(q) \tag{4.5}$$

The daily percentage change of the margin account during trading day t ($\Delta V(t)$), can now be defined as

$$\Delta V(t) = \frac{V(t) - V(t-1)}{V(t-1)} \tag{4.6}$$

In order to provide a generic, normalized, notion of the margin account's balance, it is useful to compute the cumulative sum of the daily percentage changes. The relative account balance, at the end of trading day t ($\delta(t)$) is set as the cumulative product of its daily percentage changes, as

$$\delta(t) = \begin{cases} 1 & \text{for} \quad t = 0 \\ (\Delta V(t) + 1) \times \delta(t-1) & \text{for} \quad t > 0 \end{cases} \tag{4.7}$$

This way, we can monitor the margin account's balance at the end of each trading day. $\delta(t) = 1$ indicates that the account balance, at the end of day t, is equal to the account's initial value, while $\delta(t) = 1.25$ denotes a 25% return from its opening, until the end of day t.

4.2 Strategy's Performance

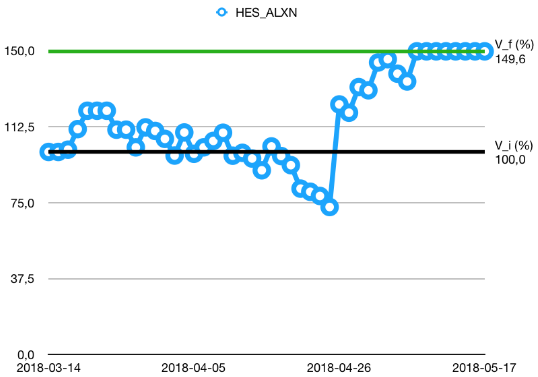
It is important to remind that, the methodology adopted in this work, presupposes that each traded pair has a specific margin account associated. In addition, the total invested capital, per trading period, is to remain constant. As a consequence, the value invested in each pair, depends on the number of pairs available for trading, at that starting date. Each pair's starting investment, at each starting date i (V_i), can then be computed as

$$V_i = \frac{I_0}{n_i} \tag{4.8}$$

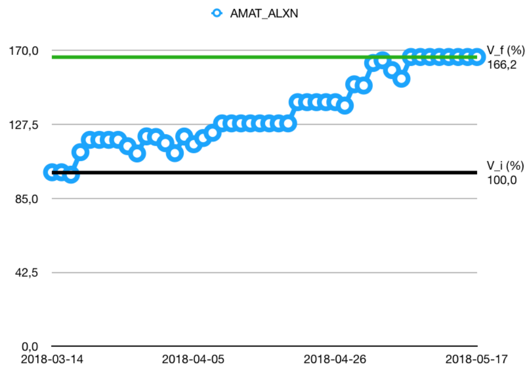
where I_0 is the available investment capital and n_i is the number of available pairs for trading at the considered starting date. I_0's value is assumed to be constant throughout the 15 different starting dates. The absolute value of I_0 is not important for our analysis (as only percentage performance interests).

As it is easily understandable, the performance of the strategy, across a given trading period, is a function of the performance of all traded pairs during that interval. In order to compute its overall performance, we first need to fix the selection method, the training window (N), and the strategy's duration (T). This procedure will be explained, in this chapter, resorting to a practical example, of a trading period, considered in this work.

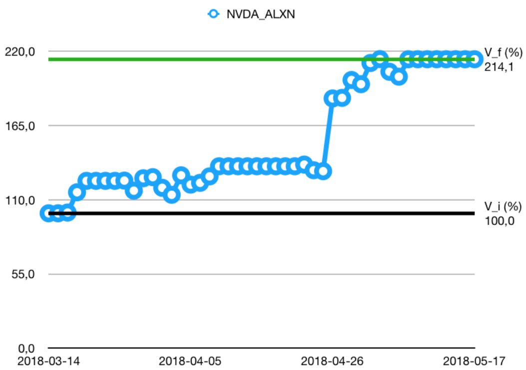
Figure 4.1, shown on the following page, portraits this example. This session took place from 2018-03-14 to 2018-05-17 ($T = 45$ business days). Stocks were selected, by the high Beta criterion, considering a training period of 100 days ($N = 100$). At this particular point in time, and using this selection method, four different pairs successfully passed the ADF filtering stage. Hence, four margin accounts were opened, each containing V_i dollars (25% of the available capital (I_0)). The evolution, in percentage, of these four accounts, throughout the 45 days, is presented on Figure 4.1's sub-figures (a), (b), (c) and (d). Details on the traded securities are available on each sub-figure's caption.



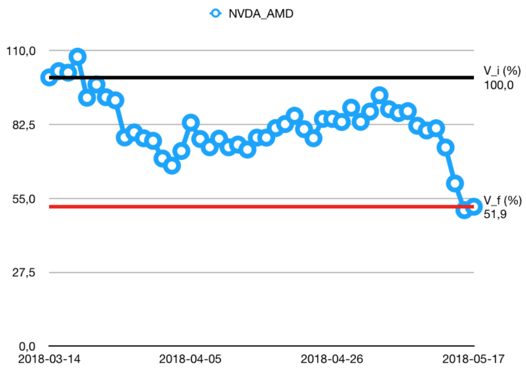
(a) Hess Corporation (HES) and Alexion Pharmaceuticals (ALXN) pair's margin account's evolution from 2018-03-14 to 2018-05-17 (blue). Initial account balance in percentage, V_i (black) and account balance at the end of the 45 days, in percentage, V_f (green).



(b) Applied Materials, Inc. (AMAT) and Alexion Pharmaceuticals (ALXN) pair's margin account's evolution from 2018-03-14 to 2018-05-17 (blue). Initial account balance in percentage, V_i (black) and account balance at the end of the 45 days, in percentage, V_f (green).



(c) Nvidia (NVDA) and Alexion Pharmaceuticals (ALXN) pair's margin account's evolution from 2018-03-14 to 2018-05-17 (blue). Initial account balance in percentage, V_i (black) and account balance at the end of the 45 days, in percentage, V_f (green).



(d) Nvidia (NVDA) and Advanced Micro Devices (AMD) pair's margin account's evolution from 2018-03-14 to 2018-05-17 (blue). Initial account balance, in percentage, V_i (black) and account balance at the end of the 45 days, in percentage, V_f (red).

Figure 4.1: Percentage evolution of the margin account associated with pair HES/ALXN (a), AMAT/ALXN (b), NVDA/ALXN (c) and NVDA/AMD (d). The test started at the end of 2018-03-14 and continued until 2018-05-17. All pairs began the test with an initial account balance of V_i dollars. Furthermore, all stocks were selected by the high beta criterion, using a $N = 100$ day training window. The final balance of each trading account is shown, on each sub-figure, and denoted as V_f.

As we can see, the four pairs demonstrated diverse behaviours. The return, in percentage, of each individual account (r_{pair}), can be computed as $V_f - V_i$. Pairs depicted on Figure 4.1 (a), (b) and (c), revealed profits of 49.6%, 66.2% and 114.1%, respectively, while the pair shown on Figure 4.1 (d), disclosed a loss of 48.1%, of the initial account balance.

Having set the balance of the accounts, at the end of the 45 days, it is now possible to quantify the performance, over that period. The evolution of our wealth can be computed, across the testing period, as the mean of all individual margin accounts. As a consequence, our wealth, at the end of trading day t ($I(t)$), can be set as

$$I(t) = (1 + \frac{\overline{r_{pair}(t)}}{100}) \times I_0 \tag{4.9}$$

where $\overline{r_{pair}(t)}$ is the average return (in percentage), of all traded margin accounts, at the end of trading day t. For the previously discussed case (Figure 4.1), at the end of the 45^{th} testing day ($t = 45$), the investor's wealth would be equal to $I(45) = 1.455 \times I_0$ (indicating a 45.5% overall profit across that period). If, at that starting date, no pair had passed the filtering ADF test, then $I(t) = I_0$.

Let us continue to consider that same example. We can compute the wealth, in terms of I_0, at the end of each trading day. Figure 4.2 pictures the investor's wealth evolution, throughout the test session.

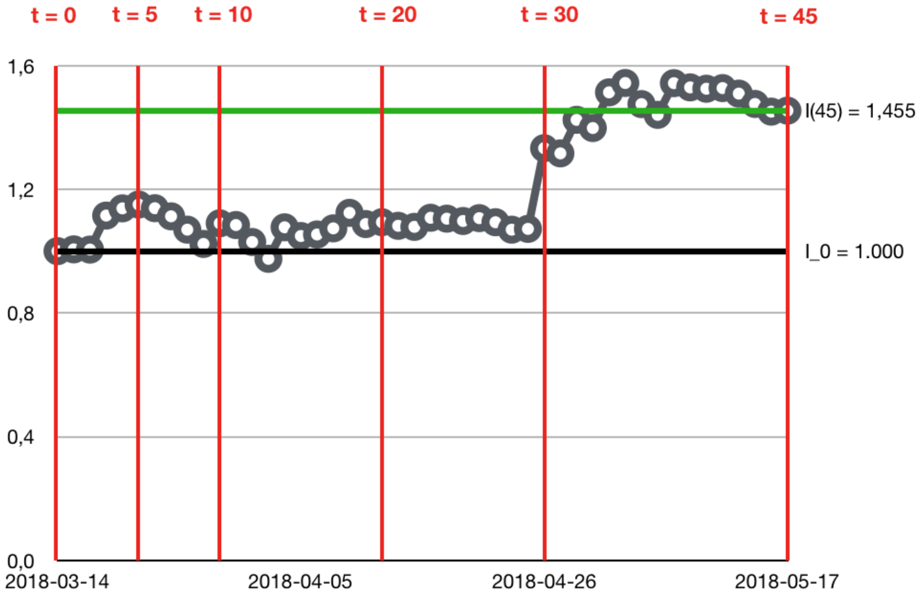


Figure 4.2: Evolution of the investor's capital, at the end of each trading day, from 2018-03-14 to 2018-05-17 (represented by blact points). Horizontal green line highlighting the overall balance at the end of the 45^{th} day. Horizontal black line representing the initial available capital (I_0). Vertical red lines highlighting the 5^{th}, 10^{th}, 20^{th}, 30^{th} and 45^{th} investment day.

As stated above, the return of the strategy, on this particular 45 day period, was 45.5%. In this dissertation, however, we will also analyze results at the end of different days. In fact, all simulation results shall be monitored at the end of the T^{th} day (5^{th}, 10^{th}, 20^{th}, 30^{th} and 45^{th} day). Basically, this procedure is the same as setting the strategy's duration. At the end of the T^{th} day, all open positions are hypothetically closed, and the results analyzed.

The returns at the end of each T^{th} day, for the case shown on Figure 4.2, are presented on Table 4.1.

T	5	10	20	30	45
Average Return (%)	15.1	9.0	9.4	33.3	45.5

Table 4.1: Percentage returns at the end of the 5^{th}, 10^{th}, 20^{th}, 30^{th} and 45^{th} day, for the case of the example aforementioned.

Nonetheless, for now, it is useful to keep our focus on the example discussed in this chapter. In other words, let us continue to examine the case in which stocks are selected, by the high beta criterion, and considering a $N = 100$ day training interval. Furthermore, let the length of the strategy remain fixed, and equal, to $T = 45$ days.

As we know, that very same procedure was followed at 15 different points in time, identified by 15 starting dates. The example discussed, so far, in this section, concerns the test started at 2018-03-14. The remaining tests, and their corresponding number of traded pairs and average returns, are presented in the following table.

Starting Date	**Number of Traded Pairs**	**Average Return During Period (%)**
2017-01-03	2	-15.6
2017-03-16	1	-99.5
2017-05-26	0	0.0
2017-08-08	0	0.0
2017-10-18	1	-28.9
2017-12-29	2	-40.6
2018-03-14	**4**	**45.5**
2018-05-24	1	-29.4
2018-08-06	0	0.0
2018-10-16	4	38.2
2018-12-28	5	3.9
2019-03-13	1	21.3
2019-05-23	0	0.0
2019-08-05	0	0.0
2019-10-15	4	19.6
Strategy Average	1.7	-5.7

Table 4.2: Number of traded pairs, and corresponding, average, percentual returns, per starting date, at the end of the $T = 45$ day trading session, for stocks selected by the high Beta criterium, using a $N = 100$ day training interval.

Table 4.2 presents the number of traded pairs, by session, for these selection method and N, settings. On it, it's also possible to observe the corresponding average return, on all traded pair's margin account, at the end of the $T = 45$ day test. In the table's bottom line, we can see the average number of pairs traded per session, and the overall average return of the strategy. Highlighted in bold, we can see the data associated with the previously discussed trading period, started 14^{th} March, 2018.

At this moment, we can rewrite Table 4.1, using the average values, across all the 15 trading sessions, as

Strategy length (days)	5	10	20	30	45
Absolute Return (%)	1.0	0.0	-5.1	-4.2	-5.7

Table 4.3: Percentage returns at the end of the 5^{th}, 10^{th}, 20^{th}, 30^{th} and 45^{th} day, for the case of the example aforementioned.

Undeniably, these particular settings do not seem to produce profitable returns. However, the magnitude of the strategy's performance, can be inappropriately quantified, by resuming it to the average return over the trading period. With that in mind, it is imperative to introduce some, indispensable, trading performance monitoring tools.

4.2.1 Hit Ratio

The concept of Hit Ratio is a fairly simple one. As we know, in our strategy, only pairs which successfully pass the ADF test undergo trading periods. The total number of filtered pairs, for all 15 sessions, per stock selection method, and per N, are shown in Table 4.4.

N	**High Beta**	**Unit Beta**	**Low Hurst**
100	25	27	73
120	28	21	55
140	40	20	73
160	39	26	43
180	31	18	52
200	23	22	65
250	13	15	69
300	14	21	66
Total	**213**	**170**	**496**

Table 4.4: Total number of filtered pairs, per stock selection method and per N. The total number of filtered pairs per selection criterion is shown in the table's last row.

These pairs' performance can be roughly divided into three main categories:

- positive,
- negative,
- neutral.

Positive performances correspond to pairs whose return at the end of the T day strategy is positive, and negative performances correspond to pairs whose return at the end of the T day strategy is negative. Neutral pairs are those which are never traded, during that time period. In other words, filtered pairs can undergo a certain test session, without opening any position.

With that in mind, we can define the Hit Ratio (HR) as

$$HR = \frac{p}{p+n} \tag{4.10}$$

where p is the number of pairs with positive returns ($r_{pair} > 0$), and n is the number of pairs with negative returns ($r_{pair} < 0$), at the end of the T day strategy.

Obviously, this indicator only serves as a measure of the ratio between positive and negative performances. Even so, it can provide us with an idea of how successfully the pairs are being chosen, and filtered.

4.2.2 Annualized returns

Annualized returns serve as a measure of return per unit of time. In this case, the considered unit of time is, quite obviously, a year. The principle, behind such methodology, is that a return of $a\%$ after 10 days, in not the same as an $a\%$ return after 10 years. In other words, the higher the return per unit of time, the better.

For our case, and recalling that the U.S. Stock Market is open around 252 day per year, the strategy's average returns per period (table 4.3), shall be extrapolated for annualized returns. These annual returns can be computed as

$$r_{annualized} = r_{period} \times \frac{252}{T} \tag{4.11}$$

where T is the strategy's duration in days and r_{period} is its average absolute return over the T day period. It is worth reminding that the methodology adopted in this dissertation, presupposes a fixed initial investment, independently from the performance of previous sessions.

That being said, we can rewrite the strategy's absolute returns, per trading period, shown in table 4.3, as

Strategy length (days)	5	10	20	30	45
Absolute return (%)	1.0	0.0	-5.1	-4.2	-5.7
Annualized return (%)	51.7	0.2	-64.2	-35	-31.9

Table 4.5: Percentage annualized returns, for 5, 10, 20, 30 and 45 day length strategies, with the high Beta selection criterion, using $N = 100$.

As it is easily understandable, the strategy's length can have a large impact on the amplification of its returns. The smaller the length, the higher the magnitude of this amplification.

Nevertheless, it is vital to compare results, taking into account the risk profile, of each strategy. The risk profile, for our case, is modeled as the standard deviation of the returns. To best understand such principles, in this chapter, we will also introduce the concept of Sharpe Ratio (Section 4.2.4). First, however, we must define a method for quantifying losses.

4.2.3 Maximum draw down

The next instrument we will discuss is known as Draw Down. This mechanism allows for a quantification of drops in a given pair's margin account, across the trading period. In fact, the Draw Down, at the end of each trading day t ($DD(t)$), can be interpreted as the percentage difference, between the maximum balance up until that point, and the current account balance. This is one of the most widely used monitoring parameters, and it is vital for any investment strategy, as traders struggle to deal with losses.

The Draw Down, can be computed, at the end of each day as

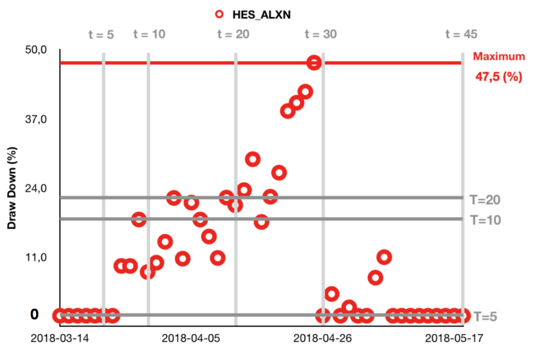
$$DD(t) = |\max_{\{i \in [W]\}} \delta_{pair}(i) - \delta_{pair}(t)|, [W] = 0, 1, ..., t \tag{4.12}$$

where $\delta_{pair}(i)$ is the designated margin account's balance at the end of day i. Evidently, a Draw Down of $DD(t) = 0$, indicates that the account, at the end of day t, is on an up to date maximum. On the other hand, a Draw Down of $DD(t) = 0,1$ indicates that, at the end of day t, the account is 10% below that maximum.

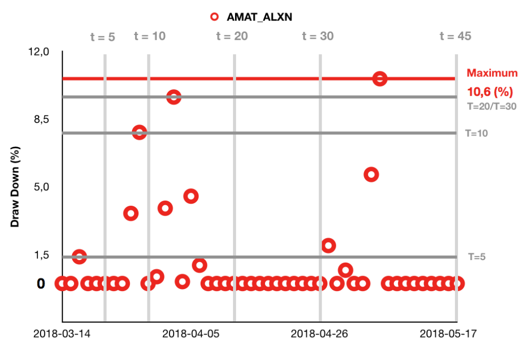
Having defined the Draw Down's value at the end of each day, we now can compute the respective Maximum Drawn Down, at that moment ($MDD(t)$). The Maximum Drawn Down, can be interpreted as the maximum absolute drop, of the account's balance, during that period, an can be set as

$$MDD(t) = \max_{\{i \in [W]\}} DD(i), [W] = 0, 1, ..., t \tag{4.13}$$

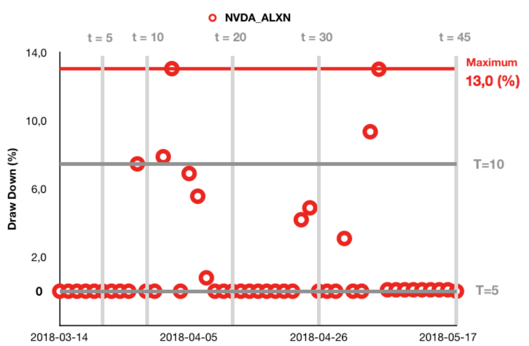
Figure 4.3, shown below, serves as a base to understand the computation procedure followed in this work. Once again, the example used so far in this chapter (Figure 4.1), shall be adopted, as illustration, for Draw Down and Maximum Draw Down computation, throughout the trading sessions. Figure 4.3's sub-figures (a), (b), (c) and (d), portrait the plots, for each pair's margin account, of these two indicators, across, the $T = 45$ day, trading period. On them, represented by red dots, we can see the values of the Draw Down at the end of each test day ($DD(t)$). Also highlighted in red, is a horizontal line, marking the Maximum Draw Down, after the 45 investment days.



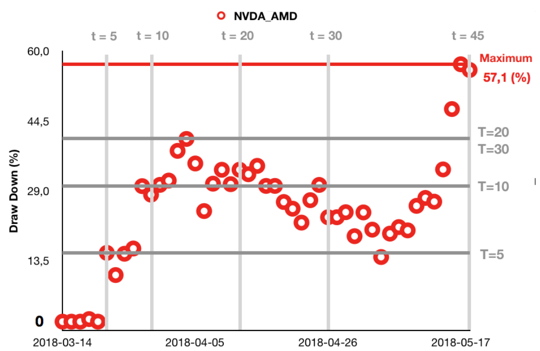
(a) Hess Corporation (HES) and Alexion Pharmaceuticals (ALXN) pair's margin account's Draw Down at the end of each day ($DD(t)$), from 2018-03-14 to 2018-05-17 (red dots).



(b) Applied Materials, Inc. (AMAT) and Alexion Pharmaceuticals (ALXN) pair's margin account's Draw Down at the end of each day ($DD(t)$), from 2018-03-14 to 2018-05-17 (red dots).



(c) Nvidia (NVDA) and Alexion Pharmaceuticals (ALXN) pair's margin account's Draw Down at the end of each day ($DD(t)$), from 2018-03-14 to 2018-05-17 (red dots).



(d) Nvidia (NVDA) and Advanced Micro Devices (AMD) pair's margin account's Draw Down at the end of each day ($DD(t)$), from 2018-03-14 to 2018-05-17 (red dots).

Figure 4.3: Draw Down (red dots), in percentage, at each day, associated with the margin account of pair HES/ALXN (a), AMAT/ALXN (b), NVDA/ALXN (c) and NVDA/AMD (d). The Maximum Draw Down, in percetage, at the end of 45 investment days, for each trading account, is represented by a red horizontal line, on each sub-figure. Light grey vertical lines signalling the 5^{th}, 10^{th}, 20^{th}, 30^{th} and 45^{th} investment days. Dark grey horizontal lines highlighting the Maximum Draw Down for a T day duration trading session. The test started at the end of 2018-03-14 and continued until 2018-05-17. Furthermore, all stocks were selected by the high beta criterion, using a $N = 100$ day training window.

As we can see, the Maximum Draw Down's value, at the end of 45 days, for the margin accounts, represented on Figures 4.3 (a), (b), (c) and (d), was, respectively, $47,5\%$, $10,6\%$, $13,0\%$ and $57,1\%$. Having computed the Maximum Draw Down, for all traded pairs' margin accounts, at the end of the 45^{th} investment day, it is possible to calculate the period's average. For this case, the test's average Maximum Draw Down was $32,0\%$.

The following Table (4.6), depicts the average Maximum Draw Down results, for the remaining tests, executed with these selection method, N and T settings. Each test is, once again, identified by its starting date.

Starting Date	Number of Traded Pairs	Average Maximum Draw Down During Period (%)
2017-01-03	2	50.1
2017-03-16	1	107.9
2017-05-26	0	0.0
2017-08-08	0	0.0
2017-10-18	1	39.5
2017-12-29	2	54.8
2018-03-14	**4**	**32.0**
2018-05-24	1	62.7
2018-08-06	0	0.0
2018-10-16	4	6.7
2018-12-28	5	37.2
2019-03-13	1	19.5
2019-05-23	0	0.0
2019-08-05	0	0.0
2019-10-15	4	26.0
Strategy Average	1.7	29.1

Table 4.6: Number of traded pairs, and corresponding, average, percentage Maximum Draw Down (in percentage), per starting date, at the end of each $T = 45$ day trading session, for stocks selected by the high Beta criterion, using a $N = 100$ day training interval. Example period highlighted in bold.

Noticeably, the strategy's average Maximum Draw Down, after 45 days of investment, across all the analyzed sessions, was 29.1%. Similarly to the previous case (average returns per period), we can also monitor the Maximum Draw Down's global average, assuming different strategy lengths. For these selection method (by high Beta) and N ($N = 100$) settings, we have

Strategy length (days)	5	10	20	30	45
Maximum Draw Down (%)	2.5	7.1	15.3	21.2	29.1

Table 4.7: Percentage Maximum Draw Down at the end of the 5^{th}, 10^{th}, 20^{th}, 30^{th} and 45^{th} day, for the strategy with the high Beta selection criterion, using $N = 100$.

As we can see, each margin account's Maximum Draw Down seems to increase, with increasing strategy length. This is a critical performance indicator, since large Maximum Draw Downs, can imply divergences on the price series of the two assets which form each pair.

All the performance monitoring mechanisms introduced in this chapter will be used, in the following chapter, to analyze the global results of this work. These global results, will take into account all three selection criteria, as well as different combinations of N and T.

4.2.4 Sharpe Ratio

The underlying concept in CAPM is that an investor can combine the investment in a risky asset (which has uncertain expected return), with an investment in a risk free product who's return rate is known and equal to r_f. That being said, it is of the out most importance to find ways of comparing investment strategies. To properly do so, we are required to relate the profit derived from a given financial product with its corresponding risk profile. Nobel laureate, William F. Sharpe, developed a formula to help investors answer these questions. This concept is known as Sharpe Ratio and it has been widely studied and debated in the world of finance [27–33].

To best understand this principle, let us consider the following image. The observed curve acts as a frontier, containing all possible asset portfolios within it. An arbitrary portfolio of risky investments is represented by the point **P**. A risk free investment with a return r_f is represented by the point **F**. The expected return, r, is presented on the y-axis and the risk, or standard deviation, σ is presented on the x-axis.

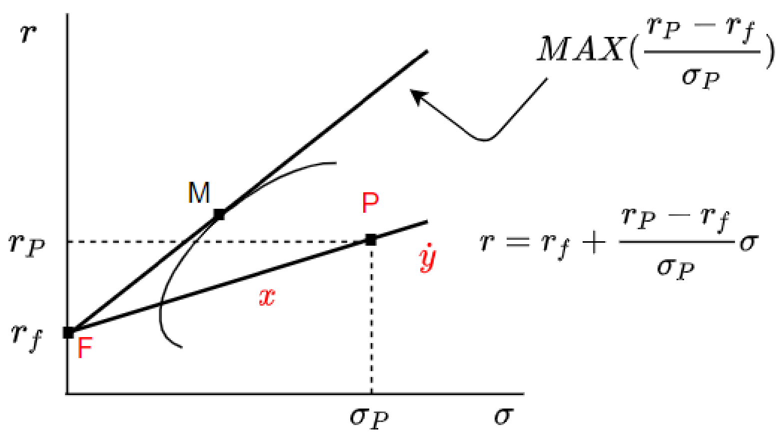


Figure 4.4: Capital Market Line

By analysing the points **F** and **P**, and by defining the line that contains both of them, we arrive at

$$r = r_f + \frac{r_p - r_f}{\sigma_P}\sigma \tag{4.14}$$

Every point on this line represents a different combination of investment in either **F** or **P**. For example, assuming that point **x**, represented on figure 4.4, corresponds to investing a fraction α_F of the total investment in **F**, and a fraction $(1 - \alpha_F)$ in **P**, we can write the following relations

$$r_x = r_f + \frac{r_p - r_f}{\sigma_P}\sigma_x \tag{4.15}$$

$$r_x = \alpha_F r_f + (1 - \alpha_F) r_P \tag{4.16}$$

Equalizing (4.15) and (4.16) hields

$$\sigma_x = (1 - \alpha_F)\sigma_P \tag{4.17}$$

When it comes to point **x** it is easy to understand that $0 < \alpha_F < 1$, yet, for point **y**, $\alpha_F < 0$. In practice, this means the investor would hold a short position in **F**, corresponding to borrowing money at the interest rate r_f.

By rewritiing (4.15)

$$\frac{r_x - r_f}{\sigma_x} = \frac{r_P - r_f}{\sigma_P} \tag{4.18}$$

where the numerator of the first term represents the increased expected return of asset **x** in relation to **F**, and the denominator represents the risk, or standard deviation, associated with an investment in **x**. In other words, the numerator is a measure of **x**'s performance, and the denominator is a measure of **x**'s uncertainty.

Most importantly, what equation (4.18) yields, is that every point on the **FP** line has the same relative performance by unit of risk. However, as an investment decision, point **M** should be considered since points located on the **FM** line present the best relative performance by unit of risk, amongst all possible portfolios. These concepts are the foundations of the Modern Portfolio Theory introduced by Harry Markowitz [34]. This theory falls out of this dissertation's scope. However, the Sharpe Ratio will be used as a measure of relative performance.

As for determining future performance, the need for comparison with a risk free investment diminishes. Any indicator of performance relative to risk can be used, in a way that best suits the investor. Nevertheless, the most commonly used approach is the quotient between annualized returns and annualized volatility of results. In fact, in this dissertation, that methodology is to be adopted.

As a consequence, we can compute the annualized Sharpe Ratio of the strategy as

$$sharpe = \frac{\overline{r_{daily}}}{\sigma_{daily}} \times \sqrt{252} \tag{4.19}$$

where $\overline{r_{daily}}$ is the average daily return and σ_{daily} is the daily standard deviation of the returns.

In our case, we start by computing the Average Daily Return (ADR), and the Average Daily Variance (ADV), for each trading session. Using, once again, the example from Figure 4.2, we can compute the ADR (of our wealth) and the ADV (also of our wealth), during that period. By repeating the process for the remaining 14 sessions, we arrive at the results shown in Table 4.8. The data associated with Figure 4.2's case, is highlighted in bold.

Starting Date (i)	$ADR_i(\%)$	ADV_i
2017-01-03 (1)	2.6×10^{-2}	8.4×10^{-3}
2017-03-16 (2)	-3.3	2.3×10^{-2}
2017-05-26 (3)	0	0
2017-08-08 (4)	0	0
2017-10-18 (5)	-0.7	1.4×10^{-3}
2017-12-29 (6)	-0.8	8.2×10^{-3}
2018-03-14 (7)	**0.8**	6.4×10^{-3}
2018-05-24 (8)	-0.4	8.9×10^{-3}
2018-08-06 (9)	0	0
2018-10-16 (10)	0.7	1.0×10^{-3}
208-12-28 (11)	0.8	2.2×10^{-2}
2018-03-13 (12)	0.5	1.1×10^{-3}
2018-05-23 (13)	0	0
2018-08-05 (14)	0	0
2019-10-15 (15)	0.4	3.7×10^{-3}
SUM	**-1.8**	8.5×10^{-2}

Table 4.8: Average Daily Return (ADR), and the Average Daily Variance (ADV), for all 15 trading sessions, for stocks selected by the high Beta criterion, with $N = 100$ and $T = 45$.

With the data displayed in Table 4.8, it is now easy to compute the overall Sharpe Ratio of the strategy (for these section method, N and T settings). The strategy's Sharpe Ratio is defined as

$$sharpe_{(strategy)} = \frac{\overline{ADR}}{\overline{\sigma_{daily}}} \times \sqrt{252} \tag{4.20}$$

where $\overline{ADR}$ is the strategy's average daily return, and $\overline{\sigma_{daily}}$ is the strategy's average daily standard deviation. The $\overline{ADR}$ and $\overline{\sigma_{daily}}$ can be computed as

$$\begin{cases} \overline{ADR} = \frac{\sum_{i=1}^{15} ADR_i}{15} \\ \overline{\sigma_{daily}} = \sqrt{\frac{\sum_{i=1}^{15} ADV_i}{15}} \end{cases} \tag{4.21}$$

For the case of the strategy shown in Table 4.8, we would have a Sharpe Ratio of $sharpe_{(strategy)} = -0.25$.

Having defined, discussed and analyzed, all the performance monitoring tools, to be used in this work, it is now possible to present, and examine, the obtained results.

Chapter 5

Results

In this chapter, we will present and analyze, the results of this work. The Maximum Draw Down, the annualized returns and the annualized Sharpe ratios, of all simulations, shall be discussed and compared. We will try to dispense valid conclusions regarding the strategy's implementation parameters. Furthermore, cointegration results, from all three stock selection methods, will also be bestowed in this segment. In fact, we will start by examining the probability of two stocks being cointegratable, as an outcome of their selection process.

5.1 Cointegrability Results

As we know, in this dissertation, three different selection methods for stocks were embraced:

- Highest Beta criterion,
- Beta closest to unit criterion,
- Lowest Hurst Exponent criterion.

At this point, it is worth remembering that 10 stocks are selected by each criterion. Three stock groups are then formed, and assets contained in each group, are paired. As a consequence, each selection method, originates 45 asset pairs. Afterwards, these pairs are tested for cointegrability (ADF Test), and only those who are deemed cointegratable, advance to the trading stage.

In this section, we will analyze a given pair's probability of being cointegratable, as a consequence of its selection process, and training interval length (N). In this work, 8 different N day length intervals were assumed for training (N equal to 100, 120, 140, 160, 180, 200, 250 and 300 days).

Table 5.1, shown below, depicts the average number of cointegratable pairs, when a N day training period is considered, for each of the stock choosing methods. The example, used throughout the previous chapter (Chapter 4), is highlighted in bold.

Training length in days (N)	High Beta	Unit Beta	Low Hurst
100	**1.7**	1.8	4.9
120	1.9	1.4	3.7
140	2.7	1.3	4.9
160	2.6	1.7	2.9
180	2.1	1.2	3.5
200	1.5	1.5	4.3
250	0.9	1.0	4.6
300	0.9	1.4	4.4

Table 5.1: Average number of cointegratable pairs, for each N, and for each selection method, across all simulated trading session.

In order to provide a probabilistic result, we can rewrite Table 5.1, in terms of percentage, of cointegratable pairs, within the 45 considered ones. By doing so, Table 5.1 takes the form

Training length in days (N)	High Beta (%)	Unit Beta (%)	Low Hurst (%)
100	**3.7**	4.0	10.8
120	4.1	3.1	8.1
140	5.9	3.0	10.8
160	5.8	3.9	6.4
180	4.6	2.7	7.7
200	3.4	3.3	9.6
250	1.9	2.2	10.2
300	2.1	3.1	9.8
Average	**3.9**	**3.1**	**9.2**

Table 5.2: Percentage of cointegratable pairs, for each N, and for each selection method, across all simulated trading sessions. The average results, for each selection criterion, are shown in the table's last row.

As we can see, pairs formed by combining stocks, selected by each of the three methods, reveal very different probabilities of being cointegratable. While the High Beta and Unit Beta criteria, denote cointegration probabilities of 3.9% and 3.1%, respectively, pairs formed using stocks with low Hurst Exponent values, have a 9.2% chance of being cointegratable. Therefore, in a preliminary analysis, it appears that the Low Hurst Exponent procedure, is a better selection process, as it is more likely to produce cointegratable pairs.

However, let us first analyze the overall behaviour of the traded pairs.

5.2 Hit Ratio

In this section, we will examine the Hit Ratio of each strategy. The Hit Ratio, helps to quantify the ratio between positive and negative return pairs, after the T day trading session, amongst those which successfully passed the filtering stage. Figures 5.1, 5.2 and 5.3, shown below, present this ratio, in percentage, after T investment days for the different training stage settings. Results are displayed resorting to a heat map, ranging from vivid red (for small values between 0 and 15%) to yellow (for values between 45 and 50%), and from light green (for values between 50 and 55%), to dark green (for values over 90%).

High Beta
T

N	5	10	20	30	45
100	55,6	58,8	66,7	69,8	77,6
120	69,2	57,9	48,0	44,4	46,2
140	33,3	42,9	59,1	54,2	45,2
160	42,9	51,4	60,4	52,8	46,9
180	61,1	56,5	50,0	50,0	44,8
200	66,7	68,4	53,8	54,8	57,7
250	66,7	56,3	50,0	56,3	66,7
300	66,7	50,0	25,0	46,2	53,8

Figure 5.1: Heat map of the number of pairs with positive returns, in percentage, for pairs formed by combining stocks selected by the high Beta criterion. Row labels, in blue, represent the training length, in days (N), and the column labels, also in blue, denote the strategy's length, in days (T). The colour scheme, ranges from vivid red (for small values between 0 and 15%) to yellow (for values between 45 and 50%), and from light green (for values between 50 and 55%), to dark green (for values over 90%).

On Figure 5.1, we can observe the Heat Ratio results, for pairs constructed using stocks with high Beta profiles. Highlighted in white, we can note the N day training intervals ($N = 100$, $N = 200$ and $N = 250$) which disclosed positive (above 50%) values, across all 5 different T day sessions.

Unit Beta
T

N	5	10	20	30	45
100	53,8	84,2	77,8	70,4	74,1
120	37,5	44,4	50,0	47,6	47,6
140	25,0	54,5	60,0	50,0	45,0
160	44,4	68,4	82,6	65,4	46,2
180	14,3	30,0	33,3	23,5	27,8
200	25,0	66,7	63,2	50,0	50,0
250	50,0	50,0	64,3	60,0	53,3
300	44,4	40,0	55,6	47,6	47,6

Figure 5.2: Heat map of the number of pairs with positive returns, in percentage, for pairs formed by combining stocks selected by the closest to unit Beta criterion. Row labels, in blue, represent the training length, in days (N), and the column labels, also in blue, denote the strategy's length, in days (T). The colour scheme, ranges from vivid red (for small values between 0 and 15%) to yellow (for values between 45 and 50%), and from light green (for values between 50 and 55%), to dark green (for values over 90%).

As it is easily noticeable, the strategies revealed similar Hit Ratio patterns. In fact, when we compare the two market based selection methods, we can remark that, once again, for $N = 100$ and $N = 250$ (highlighted in white), positive results were achieved for every different strategy length. Furthermore, for

$N = 200$, with the exception of the first, $T = 5$ day interval, all remaining sessions revealed Hit Ratios above 50%.

Low Hurst
T

N	5	10	20	30	45
100	34,6	64,8	62,7	62,5	63,0
120	82,6	63,8	76,5	72,7	54,5
140	92,6	82,0	77,8	66,2	60,3
160	50,0	56,0	67,6	55,8	62,8
180	59,1	51,2	59,6	59,6	61,5
200	58,6	55,6	59,7	65,6	49,2
250	40,9	48,9	43,9	46,4	53,6
300	33,3	33,3	37,7	40,6	40,9

Figure 5.3: Heat map of the number of pairs with positive returns, in percentage, for pairs formed by combining stocks selected by the low Hurst criterion. Row labels, in blue, represent the training length, in days (N), and the column labels, also in blue, denote the strategy's length, in days (T). The colour scheme, ranges from vivid red (for small values between 0 and 15%) to yellow (for values between 45 and 50%), and from light green (for values between 50 and 55%), to dark green (for values over 90%).

Finally, the low Hurst criterion appears to reveal the best results. In fact, for this selection method, the great majority of the strategies denoted Hit Ratios above 50%. Highlighted in white, we can see the training settings, which led to positive Hit Ratios, throughout all strategy lengths ($N = 120$, $N = 140$, $N = 160$ and $N = 180$). It is also interesting to note that, for the $N = 100$ and $N = 200$ strategies, all but one lengths, denoted values over 50%.

Despite being clearly a good sign, the Hit Ratio can be quite misleading, as the magnitude of gains, and losses, is not being considered. As we shall see, this indicator can lead us in a rather rashed interpretation of results. Therefore, it is also important to analyze the remaining performance tools at our disposal, attempting to provide a more realistic picture of the strategies' performance.

5.3 Absolute Returns per Period

In order to properly analyze our results, it is imperative to examine the average absolute returns, per session. This indicator, acts as a precise measure, of how much, the investor would have gained, or lost, in percentage, per trading period, if he had invested in such strategy. The average results (across all 15 simulations), for each strategy, are presented on Figures 5.4, 5.5 and 5.6. The example strategy, presented in Table 4.2, which was used for illustrating the computation of such results, is highlighted, by a white circle, on Figure 5.4.

High Beta

N \ T	5	10	20	30	45
100	1,0	0,0	-5,1	-4,2	-5,7
120	-0,1	-4,8	-4,9	-6,6	-9,9
140	-1,2	-3,3	3,8	4,0	-3,6
160	0,4	-8,1	-8,5	-13,0	-19,9
180	-0,3	-0,1	-5,4	-5,4	-5,9
200	0,3	-1,3	-12,6	-16,0	-13,7
250	0,5	-6,8	-9,3	-4,3	-3,1
300	-0,3	-4,0	-7,5	-5,2	-5,5

Figure 5.4: Heat map of the average absolute return, per T day session, in percentage, for pairs formed by combining stocks selected by the high Beta criterion. Row labels, in blue, represent the training length, in days (N), and the column labels, also in blue, denote the strategy's length, in days (T). The colour scheme, ranges from yellow (for small losses between 0 and 0.5%), to vivid red (for losses over 10%), and from light green (for small profits between 0 and 0.5%) to dark green (for profits over 7.5%). The example strategy, from Chapter 4 (Table 4.2), is highlighted by a white circle.

When we analyze the absolute return, for stocks whit high Beta values, it becomes clear that the results scattered. It is not possible to detect any clear return pattern and it does not seem to provide any proper investment edge.

Unit Beta

N \ T	5	10	20	30	45
100	-0,2	1,6	1,9	5,2	6,7
120	-0,2	-0,2	0,0	-0,8	-2,1
140	-1,0	-1,1	-3,7	-6,3	-8,6
160	-0,8	2,3	3,2	4,0	2,0
180	-0,3	-1,9	-2,7	-3,4	-3,4
200	-0,3	1,6	2,0	-0,6	-4,6
250	0,4	1,0	-1,1	-1,6	-4,0
300	-2,1	-2,7	-3,4	-5,8	-4,7

Figure 5.5: Heat map of the average absolute return, per T day session, in percentage, for pairs formed by combining stocks selected by the unit Beta criterion. Row labels, in blue, represent the training length, in days (N), and the column labels, also in blue, denote the strategy's length, in days (T).The colour scheme, ranges from yellow (for small losses between 0 and 0.5%), to vivid red (for losses over 10%), and from light green (for small profits between 0 and 0.5%) to dark green (for profits over 7.5%).

As we can see, stocks selected by the closest to unit Beta criterion, reveal better results than the ones selected by the high Beta method. There are clearly more strategies with positive returns on Figure 5.5 than there are on Figure 5.4. Furthermore, the number of strategies whose losses, in average, surpassed 7.5%, is much higher on Figure 5.4's case. Highlighted in white (on Figure 5.5), are two cases which merit special emphasis. You'll notice that, for $N = 100$ and $N = 160$, with the exception of a small loss during the first five investment days, positive returns were achieved in all remaining intervals.

Low Hurst

N \ T	5	10	20	30	45
100	0,4	1,8	-0,5	-0,5	-0,9
120	5,7	4,9	7,5	8,4	4,4
140	0,8	3,5	1,7	2,9	0,4
160	-1,3	-1,3	0,2	0,4	0,4
180	-0,8	0,0	1,4	1,4	3,0
200	0,6	-0,0	-0,5	-0,3	-0,5
250	-0,6	-1,1	-0,3	-2,2	0,6
300	-2,2	-6,2	-7,5	-10,5	-9,3

Figure 5.6: Heat map of the average absolute return, per T day session, in percentage, for pairs formed by combining stocks selected by the low Hurst criterion. Row labels, in blue, represent the training length, in days (N), and the column labels, also in blue, denote the strategy's length, in days (T).The colour scheme, ranges from yellow (for small losses between 0 and 0.5%), to vivid red (for losses over 10%), and from light green (for small profits between 0 and 0.5%) to dark green (for profits over 7.5%).

The third and last selection method, once again, reveals the best overall results, in comparison to the two previous cases. In fact, it is clear that strategies applied on stocks, chosen for denoting low Hurst Exponent values, have a much higher probability of being profitable. Furthermore, strategies with $N = 120$ and $N = 140$ (highlighted in white), revealed profitable performances across all strategy lengths. Also highlighted in white, it is possible to observe the absolute returns for $N = 180$. For this particular setting, we can denote a small loss (0.8%) at the end of the first five day session, followed by a consistent profit increase until the 45^{th} day.

Nevertheless, as previously stated, it can be rash to focus our analysis sole on this indicator. That being said, let us continue our examination by quantifying the, average, absolute drops in wealth, for each strategy.

5.4 Maximum Draw Down

Let us start by scrutinizing the Maximum Draw Down values for all the simulated strategies. As we know, this indicator allows for a quantification of the largest, average, absolute drop in wealth, across the T day length strategies. Maximum Draw Down results shall be presented in three segments, one for each selection method (Figures 5.7, 5.8 and 5.9).

Results for stocks selected by depicting the highest Beta values, are shown on Figure 5.7. Once again, the example strategy period, used throughout Chapter 4 to explain the Maximum Draw Down's computation (Table 4.6), is highlighted on this figure by a white circle.

High Beta

N \ T	5	10	20	30	45
100	2,5	7,1	15,3	21,2	29,1
120	2,1	9,8	15,3	20,2	28,3
140	2,8	9,7	12,6	17,3	27,0
160	1,0	12,4	19,6	28,7	34,6
180	1,4	5,6	13,8	20,9	26,3
200	1,0	7,0	21,3	29,9	34,3
250	1,2	10,4	18,2	21,9	23,7
300	0,5	6,2	12,0	15,3	20,1

Figure 5.7: Heat map of the average Maximum Draw Down, in percentage, for pairs formed by combining stocks selected by the high Beta criterion. Row labels, in blue, represent the training length, in days (N), and the column labels, also in blue, denote the strategy's length, in days (T). The colour scheme, ranges from yellow (for small values between 0 and 2.5%), to red (for values over 20%).

Clearly, for these training settings, the Maximum Draw Down values increase, with increasing strategy lengths. Furthermore, this performance indicator, appears to have an approximately linear evolution in relation to T. In fact, after 20 investment days, all strategies bestow average Maximum Draw Downs above 10%. After 45 days, all strategies, regardless of N, denote average drops in wealth, superior to 20%.

As previously stated, the same analysis was performed for the stocks whose Beta was closest to unit. The obtained results are shown below, on Figure 5.8.

Unit Beta

N \ T	5	10	20	30	45
100	0,7	1,4	6,9	11,0	13,3
120	0,5	2,2	3,9	5,6	52,2
140	2,3	5,0	10,4	17,8	129,1
160	1,7	2,7	5,3	8,1	74,4
180	1,0	3,0	7,2	10,7	86,8
200	1,0	2,4	5,4	9,4	93,6
250	1,4	2,6	6,0	8,9	82,1
300	2,5	5,2	10,7	14,8	101,1

Figure 5.8: Heat map of the average Maximum Draw Down, in percentage, for pairs formed by combining stocks selected by the unit Beta criterion. Row labels, in blue, represent the training length, in days (N), and the column labels, also in blue, denote the strategy's length, in days (T).The colour scheme, ranges from yellow (for small values between 0 and 2.5%), to red (for values over 20%).

For this selection method, the Maximum Draw Downs' values, throughout the first 30 investment days, appear to be preferable to the previous ones. Their evolution, up until that moment, looks relatively less accentuated, and results seem fairly better when compared to those derived from the high Beta criterion.

However, when we continue to increase the strategies' length ($T = 45$), Maximum Draw Down values depict a rapid escalate. In fact, in average, when set side by side with the corresponding values from Figure 5.7 ($T = 45$), they pull largely ahead.

Lastly, let us analyze the results, accomplished when the low Hurst exponent selection method was adopted.

Low Hurst

N \ T	5	10	20	30	45
100	1,7	4,9	11,4	18,7	25,2
120	1,2	5,4	8,4	13,0	21,0
140	1,9	4,4	9,9	14,9	22,1
160	3,1	4,4	8,1	12,8	20,0
180	3,7	7,1	12,0	16,1	21,3
200	1,6	5,7	13,5	17,6	24,5
250	1,6	5,2	11,9	16,3	22,3
300	3,2	9,0	17,1	21,5	27,8

Figure 5.9: Heat map of the average Maximum Draw Down, in percentage, for pairs formed by combining stocks selected by the low Hurst criterion. Row labels, in blue, represent the training length, in days (N), and the column labels, also in blue, denote the strategy's length, in days (T).The colour scheme, ranges from yellow (for small values between 0 and 2.5%), to red (for values over 20%).

As we can see, the Maximum Draw Down values, portraited on Figure 5.9, denote an evolution which is very similar to the one seen on the heat map, associated with the pairs formed by combining stocks with high Beta values (Figure 5.7). Once again, the strategies' losses, appear to increase rather constantly, with increasing strategy lengths. As for the high Beta case, at the end of the 45^{th} investment day, all strategies experienced average drops in wealth, superior to 20%.

Understandably, in an ideal scenario, it is not desirable for strategies to suffer significant losses across the trading sessions. However, when this happens, and a weighty average Maximum Draw Down exists, it serves as an indication that our wealth has suffered, a noteworthy, negative evolution. Hence, and keeping in mind the conclusion presented in this section, it appears that our approach struggles to manage losses, for larger trading periods. As a consequence, in this work, we will only analyze results for relatively small investment intervals, up to 45 days.

That being said, and having quantified the strategies' absolute losses, let us examine the annualized returns, for these, short term, investment cases.

5.5 Annualized Returns

In this section, we will present the results of the returns per year, achieved by each strategy. Quite obviously, these results will be qualitatively similar to the ones shown in section 5.3 (in terms of being positive or negative). However, when we annualize the gains, and losses, their magnitude can suffer severe alterations. This fact can be especially true, when we consider the shorter term strategies.

Figure 5.10, shown below, portraits the results obtained for high Beta stocks. On it, once again, the example strategy of Chapter 4 (Table 4.5), is highlighted, by a white circle.

High Beta

N \ T	5	10	20	30	45
100	51,7	0,2	-64,2	-35,0	-31,9
120	-5,8	-120,5	-62,2	-55,3	-55,5
140	-60,4	-83,6	48,3	33,6	-20,1
160	18,2	-204,4	-107,0	-109,5	-111,7
180	-14,7	-3,2	-67,5	-45,5	-32,8
200	13,2	-33,3	-159,3	-134,2	-76,6
250	26,2	-172,4	-116,8	-36,4	-17,2
300	-14,6	-100,4	-93,9	-43,9	-30,6

Figure 5.10: Heat map of the average annualized return, in percentage, for pairs formed by combining stocks selected by the high Beta criterion. Row labels, in blue, represent the training length, in days (N), and the column labels, also in blue, denote the strategy's length, in days (T). The colour scheme, ranges from yellow (for small annualized losses between 0 and 5%), to vivid red (for annualized losses of over 100%), and from light green (for small annualized profits between 0 and 5%) to dark green (for annualized profits of over 100%).

Without surprise, analogously to the conclusions of section 5.3, this method does not appear to produce viable results. Even so, it can be interesting to note that the best behaviours were achieved for the $T = 5$ day trading strategies. However, at the end of 10^{th} day, the returns diminish, and almost all strategies reveal losses in wealth. As a consequence, it would be unwise for an investors to apply these strategies.

That being said, let us continue to examine the results derived from the two remaining stock selection criteria. The annualized returns for the closest to unit beta selection method, are presented on the following image.

Unit Beta

N \ T	5	10	20	30	45
100	-8,0	39,6	23,5	43,8	37,8
120	-10,8	-5,6	0,3	-7,0	-11,8
140	-50,3	-28,4	-46,5	-52,7	-48,0
160	-42,5	57,7	40,2	33,9	11,4
180	-17,4	-47,7	-33,6	-28,9	-19,1
200	-16,2	41,0	25,3	-5,0	-26,0
250	21,0	24,9	-14,3	-13,7	-22,5
300	-108,3	-68,3	-42,7	-48,6	-26,3

Figure 5.11: Heat map of the average annualized return, in percentage, for pairs formed by combining stocks selected by the closest to unit Beta criterion. Row labels, in blue, represent the training length, in days (N), and the column labels, also in blue, denote the strategy's length, in days (T). The colour scheme, ranges from yellow (for small annualized losses between 0 and 5%), to vivid red (for annualized losses of over 100%), and from light green (for small annualized profits between 0 and 5%) to dark green (for annualized profits of over 100%).

As previously observed, in section 5.3, for $N = 100$ and $N = 160$ (highlighted in white), for stocks selected by the high Beta criterion, the strategies are profitable in all intervals, except for the first one ($T = 5$). For some small strategy lengths, strategies with $N = 200$ and $N = 250$, reveal positive returns. However, these cases lack consistency, as the strategies are unable to sustain positive performances throughout the remaining intervals.

Low Hurst

N \ T	5	10	20	30	45
100	18,3	45,3	-5,9	-4,1	-5,1
120	287,8	122,5	94,8	70,4	24,5
140	41,4	87,2	21,8	24,5	2,3
160	-67,1	-33,4	2,6	3,7	2,4
180	-41,1	0,2	17,7	11,8	16,6
200	32,0	-1,1	-6,9	-2,5	-2,8
250	-27,9	-27,6	-4,2	-18,4	3,4
300	-108,5	-156,9	-94,4	-88,0	-52,1

Figure 5.12: Heat map of the average annualized return, in percentage, for pairs formed by combining stocks selected by the low Hurst Exponent criterion. Row labels, in blue, represent the training length, in days (N), and the column labels, also in blue, denote the strategy's length, in days (T). The colour scheme, ranges from yellow (for small annualized losses between 0 and 5%), to vivid red (for annualized losses of over 100%), and from light green (for small annualized profits between 0 and 5%) to dark green (for annualized profits of over 100%).

Finally, on Figure 5.12, presented above, we can see the annualized returns for the last stock selection procedure (the low Hurst criterion). Once again, it is clear that this is the method which produced the best overall results. Similarly to previous cases, the most interesting results are highlighted, on the figure, by a white frame. As formerly discussed (in section 5.3), for $N = 120$ and $N = 140$, this selection method achieved positive returns for all different strategy lengths. However, when we consider the annualized returns, it becomes clear that these strategies denote higher wealth growth rates for smaller length trading sessions ($T = 5$ and $T = 10$). It is also interesting to note that, for $N = 100$, the strategy is profitable for these same trading lengths. Nevertheless, for this particular training settings, the strategy struggles to maintain performance for larger time periods.

Having presented, and examined, the annualized returns for all simulations, it is now time to analyze these results, taking also into account, the risk profile of each strategy.

5.6 Sharpe Ratio

As we know, the Sharpe Ratio serves as measure of the quotient between annualized returns, and the annualized volatility of returns. Figures 5.13, 5.14 and 5.15, depict the results obtained for the three considered stock selection procedures. Figure 5.13, shown below, concerns the high Beta procedure. On it, as for previous cases, the computation example used in section 4.2.4 (Table 4.8), is highlighted by a white circle.

High Beta

T

N	5	10	20	30	45
100	0,82	-0,01	-0,50	-0,16	-0,25
120	-0,07	-1,71	-0,32	-0,34	-0,33
140	-1,23	-1,05	0,58	0,43	-0,27
160	0,47	-1,89	-0,55	-0,81	-0,60
180	-0,35	0,03	-0,64	-0,39	-0,12
200	0,46	-0,38	-1,38	-0,69	-0,46
250	0,81	-1,91	-0,72	-0,15	-0,65
300	-0,95	-1,43	-0,76	-0,41	-0,75

Figure 5.13: Heat map of the average Sharpe Ratio, for each strategy, for pairs formed by combining stocks selected by the high Beta criterion. Row labels, in blue, represent the training length, in days (N), and the column labels, also in blue, denote the strategy's length, in days (T). The colour scheme, ranges from yellow (for small annualized losses between -0.1 and 0), to vivid red (for Sharpe Ratios under -1), and from light green (for small values between 0 and 0.1) to dark green (for values over 1.25).

As it is observable, this method leads to poor overall results. In fact, when this selection procedure was adopted, no strategy was able to produce a Sharpe Ratio superior to one. In other words, none of these strategies revealed positive annualized returns, greater than its own annualized standard deviation.

Unit Beta

N \ T	5	10	20	30	45
100	-0,39	1,31	0,48	0,88	0,88
120	-0,98	-0,15	0,05	-0,18	-0,28
140	-1,28	-0,62	-1,16	-1,08	-1,04
160	-1,00	0,84	0,77	0,70	0,26
180	-0,73	-1,96	-0,69	-0,56	-0,17
200	-0,58	0,83	0,71	-0,05	-0,50
250	0,50	0,64	-0,53	-0,59	-0,48
300	-3,67	-2,38	-1,03	-0,82	-0,62

Figure 5.14: Heat map of the average Sharpe Ratio, for each strategy, for pairs formed by combining stocks selected by the closest to unit Beta criterion. Row labels, in blue, represent the training length, in days (N), and the column labels, also in blue, denote the strategy's length, in days (T). The colour scheme, ranges from yellow (for small annualized losses between -0.1 and 0), to vivid red (for Sharpe Ratios under -1), and from light green (for small values between 0 and 0.1) to dark green (for values over 1.25).

Figure 5.14, shown above, portraits the Sharpe Ratio values for stocks selected using the closest to unit Beta criterion. As we can see, when we examine this indicator, it becomes clear that the strategies' results are more consistent for smaller T day intervals. In fact, for the two previously discussed cases ($N = 100$ and $N = 160$), the strategies' Sharpe Ratio is at its peak for $T = 10$. These two cases are, once again, highlighted on the figure by white frames. It is also interesting to not that, for $N = 100$ and $T = 10$, the strategy denoted a Sharpe Ratio superior to 1.

Low Hurst

N \ T	5	10	20	30	45
100	0,37	0,88	-0,18	-0,46	-0,25
120	2,03	1,20	1,14	0,94	0,28
140	0,48	1,35	0,39	0,45	0,11
160	-1,17	-0,68	0,10	0,14	0,13
180	-0,55	-0,03	0,31	0,25	0,43
200	0,85	0,02	0,08	0,09	0,12
250	-0,91	-0,52	0,07	-0,25	0,10
300	-2,00	-2,92	-1,29	-0,44	-0,37

Figure 5.15: Heat map of the average Sharpe Ratio, for each strategy, for pairs formed by combining stocks selected by the low Hurst Exponent criterion. Row labels, in blue, represent the training length, in days (N), and the column labels, also in blue, denote the strategy's length, in days (T). The colour scheme, ranges from yellow (for small annualized losses between -0.1 and 0), to vivid red (for Sharpe Ratios under -1), and from light green (for small values between 0 and 0.1) to dark green (for values over 1.25).

Finally, on Figure 5.15, presented above, we can see the Sharpe Ratio results for strategies in which the pairs were formed resorting to stocks with low Hurst Exponent values. As we can once again observe, this is undoubtedly the method with the best overall results. Highlighted by a white frame, are the cases already discussed in previous sections. As it is perceivable, for $N = 100$, $N = 120$ and $N = 140$, the highest Sharpe Ratio values were achieved for small investment periods (either for $T = 5$ or $T = 10$). In fact, for the training settings which led to the best results ($N = 120$), the Sharpe Ratio is at its peak for $T = 5$, and gradually diminishes across all remaining strategy lengths. Furthermore, the best N/T settings ($N = 120$ and $T = 5$), achieved a Sharpe Ratio of 2.03. This fact implies that the annualized return of this strategy, exceeded its own annualized standard deviation by a ratio of two (during the 5 investment days). It is also important to refer that, for these training settings, the strategy was able to maintain a Sharpe Ratio above one, throughout the first 20 investment days (and very close to one, for $T = 30$).

In this chapter, we have presented, examined and discussed, the results derived from all the simulations performed in this work. Having done so, we now possess the necessary information to retrieve valuable guidelines and conclusions, regarding the implementation of these short term pairs trading strategies. Such deductions will be exhibited in the following segment of this thesis (Chapter 6).

Chapter 6

Conclusions

As referenced multiple time throughout this dissertation, we set out with the objective of developing functional and applicable procedures, for implementing short term statistical arbitrage pairs trading strategies. With that in mind, we have adopted three different asset selection methodologies, and developed algorithms, to test and compare them. In addition, several important, time variant parameters, were studied and analyzed, for each selection setting. All things considered, there are several, worth noting conclusions, to retrieve from this work.

The first major verdict to be proclaimed is related to the stock selection procedures. In chapter 3, we provided a detailed description of the three different methodologies adopted throughout this dissertation. Two were based on the CAPM, and formed pairs by combining assets with either high or close to unit Beta values. In the third one, stocks with low Hurst Exponent values were selected and matched. It became clear that, not only are these methodologies vital for a practical implementation of the strategies, but they are also key in ensuring the quality of their results. Furthermore, as the same ADF filtering test was used for all strategies, significant differences regarding each of the three considered methods' performance, were perfectly observable.

In fact, when we compare the two CAPM based selection procedures, we can perceive a clear advantage in adopting the closest to unit Beta criterion. This can be explained by the fact that high Beta stocks denote higher covariances in relation to the market. Hence, those assets tend to also reveal higher covariance values between themselves. As previously mentioned (section 3.3.1), this can have a negative impact on the pair's variance, thus diminishing investment opportunity perspectives. Such impact may be even more decisive, due to the short term character of our simulations. Another possible cause for this performance difference, is related with the volatility of both stock' groups. High Beta stocks tend to present higher volatility patterns and, therefore, their prices are more likely to diverge, leading to severe losses in pairs formed with them.

Finally, and more importantly, when we compare the global performance of the three selection methods, we can observe an undeniable improvement of results, for the low Hurst Exponent criterion. In fact, this method revealed the greatest number of positive return strategies, as well as the best Hit and Sharpe ratio results. Furthermore, it became clear that pairs constructed by combining stocks with low

Hurst Exponent values, are much more likely to be cointegratable, than those derived from joining together stocks with similar Beta values. As a consequence, from this point onwards, we shall center our verdicts on the results from this particular stock selection procedure.

As we know, in chapter 5, we presented the results from different performance indicators, obtained for diverse training and trading length combinations. Once again, the objective was to develop practical and functional guidelines for real life investment strategies, and once more, important assumptions can be taken.

When it comes to the influence of the considered training length, it became clear that, for some specific configurations ($N = 120$ and $N = 140$), the strategy was able to maintain its profitability across all considered testing durations. Moreover, smaller training lengths appear to be unable to sustain profitability, for more than ten investment days. Lastly, the use of larger selection and filtering time frames, requires the implementation of longer term strategies (since these settings seem to take more days to achieve profitable results).

Finally, let us discuss the conclusions to retrieve from the impact, on results, of the five different considered session lengths. Once again, there are several worth noting guidelines to fetch from our studies. In fact, results on this subject couldn't have been clearer. For the best training settings ($N = 120$ and $N = 140$), the best returns, per unit of time, were undoubtedly obtained for shorter trading periods (either $T = 5$ or $T = 10$). Furthermore, the best Sharpe and Hit Ratios were also achieved for these T day investment sessions. As a consequence, it is fair to state that, investors should strive to implement strategies with similar trading periods.

Nonetheless, and despite our valuable deductions, there are also aspects pending future investigation, regarding implementations, and possible investments, in more realistic settings.

6.1 Future Work

The strategy presented in this work was a fairly simple and basic one. For instance, from an investor's point of view, our strategy lacks a proper loss management mechanism. In addition, several important parameters and coefficients were set a priori, and their evolution neglected (namely the weight coefficient, B). However, these facts can have a positive impact in identifying cause and effect situations. In other words, our selection and filtering methodologies' performance, can be more easily linked to our results.

That being said, there are several possible areas to improve and enhance this strategy's performance. Mainly, and as previously referred, an effective stop loss procedure must be implemented in order to deal with diverging pairs. In addition, the impact from the bid-ask spread and from broker's fees, should be considered and examined.

At last, and considering this thesis' results, a possible investor should definitely attempt to investigate to what extent it is possible to reduce the strategy's "refresh" interval. As we know, in this work, the strategy was tested at different points in time, located fifty business days apart. However, as formerly discussed, the best results were obtained for smaller time frames ($T = 5$ and $T = 10$). Hence, being able to reduce the time between sessions, without the diminishing of results, is key in improving performance and profitability.

Bibliography

[1] W. F. Sharpe. Capital asset prices: A theory of market equilibrium under conditions of risk. *The journal of finance*, 19(3):425–442, 1964.

[2] B. Do and R. Faff. Does simple pairs trading still work? *Financial Analysts Journal*, 66(4):83–95, 2010.

[3] G. Vidyamurthy. *Pairs Trading: quantitative methods and analysis*, volume 217. John Wiley & Sons, 2004.

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

[5] N. Huck and K. Afawubo. Pairs trading and selection methods: is cointegration superior? *Applied Economics*, 47(6):599–613, 2015.

[6] S. Greco, J. Figueira, and M. Ehrgott. *Multiple criteria decision analysis*, volume III. Springer, 2016.

[7] R. F. Engle and C. W. Granger. Co-integration and error correction: representation, estimation, and testing. *Econometrica: journal of the Econometric Society*, pages 251–276, 1987.

[8] D. A. Dickey and W. A. Fuller. Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American statistical association*, 74(366a):427–431, 1979.

[9] G. Elliott, T. J. Rothenberg, and J. H. Stock. Efficient tests for an autoregressive unit root. Technical report, National Bureau of Economic Research, 1992.

[10] D. Kwiatkowski, P. C. Phillips, P. Schmidt, and Y. Shin. Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root? *Journal of econometrics*, 54(1-3):159–178, 1992.

[11] P. C. Phillips and P. Perron. Testing for a unit root in time series regression. *Biometrika*, 75(2): 335–346, 1988.

[12] M. Arltova and D. Fedorova. Selection of unit root test on the basis of length of the time series and value of ar (1) parameter. *Statistika-Statistics and Economy Journal*, 96(3):47–64, 2016.

[13] S. E. Said and D. A. Dickey. Testing for unit roots in autoregressive-moving average models of unknown order. *Biometrika*, 71(3):599–607, 1984.

[14] J. M. ”Vanden. ”options trading and the capm”. *”The Review of Financial Studies”*, ”17”(”1”):”207–238”, ”2004”.

[15] P. W. Santosa and H. Y. Laksana. Value at risk, market risk and trading activity: Capm alternative model. *Journal of Applied Finance and Banking*, 1(4):239, 2011.

[16] C. Chiarella, R. Dieci, X.-Z. He, and K. Li. An evolutionary capm under heterogeneous beliefs. *Annals of Finance*, 9(2):185–215, 2013.

[17] P. Bossaerts and C. Plott. The capm in thin experimental financial markets. *Journal of Economic Dynamics and Control*, 26(7-8):1093–1112, 2002.

[18] A. Ang and J. Chen. Capm over the long run: 1926–2001. *Journal of Empirical Finance*, 14(1): 1–40, 2007.

[19] M. R. Reinganum. A new empirical perspective on the capm. *Journal of financial and quantitative analysis*, pages 439–462, 1981.

[20] J. Lewellen and S. Nagel. The conditional capm does not explain asset-pricing anomalies. *Journal of financial economics*, 82(2):289–314, 2006.

[21] H. E. Hurst. Long term storage capacity of reservoirs. *Trans. Amer. Soc. Civil Eng*, 116:770–799, 1951.

[22] J. Alvarez-Ramirez, J. Alvarez, E. Rodriguez, and G. Fernandez-Anaya. Time-varying hurst exponent for us stock markets. *Physica A: statistical mechanics and its applications*, 387(24):6159–6169, 2008.

[23] B. Qian and K. Rasheed. Hurst exponent and financial market predictability. In *IASTED conference on Financial Engineering and Applications*, pages 203–209, 2004.

[24] M. S. Raimundo and J. Okamoto Jr. Application of hurst exponent (h) and the r/s analysis in the classification of forex securities. *Int. J. Model. Optim*, 8:116–124, 2018.

[25] J. P. Ramos-Requena, J. Trinidad-Segovia, and M. Sánchez-Granero. Introducing hurst exponent in pair trading. *Physica A: statistical mechanics and its applications*, 488:39–45, 2017.

[26] M. Couillard and M. Davison. A comment on measuring the hurst exponent of financial time series. *Physica A: Statistical Mechanics and its Applications*, 348:404–418, 2005.

[27] A. W. Lo. The statistics of sharpe ratios. *Financial analysts journal*, 58(4):36–52, 2002.

[28] C. L. Israelsen et al. A refinement to the sharpe ratio and information ratio. *Journal of Asset Management*, 5(6):423–427, 2005.

[29] C. Memmel. Performance hypothesis testing with the sharpe ratio. *Available at SSRN 412588*, 2003.

[30] W. Goetzmann, J. Ingersoll, M. I. Spiegel, and I. Welch. Sharpening sharpe ratios. Technical report, National bureau of economic research, 2002.

[31] O. Ledoit and M. Wolf. Robust performance hypothesis testing with the sharpe ratio. *Journal of Empirical Finance*, 15(5):850–859, 2008.

[32] K. Dowd. Adjusting for risk:: An improved sharpe ratio. *International review of economics & finance*, 9(3):209–222, 2000.

[33] D. H. Bailey and M. Lopez de Prado. The sharpe ratio efficient frontier. *Journal of Risk*, 15(2):13, 2012.

[34] H. Markowitz. Portfolio selection. *The Journal of Finance*, 7(1):77–91, 1952. ISSN 00221082, 15406261. URL `http://www.jstor.org/stable/2975974`.