Developing short term algorithmic pairs trading strategies using time series modeling

Afonso Villaret Frischknecht
afonsofris@hotmail.com

Instituto Superior Técnico, Universidade de Lisboa, Lisboa, Portugal

July 2021

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.

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

1. Introduction

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.

For instance, [1] 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.

On the other hand, [2] 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 2.2), and stocks with low Hurst Exponent values (Section 2.3).

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 [3]. 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, researchers 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 third and last filtering method is central to this dissertation's scope. Cointegration was first proposed in 1987, by Nobel laureates [4]. 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

$$u_t = y_t - a x_t \tag{1}$$

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 [5] in 1979 and has been extensively used in time series modeling ever since. In fact, the conclusions of [3] 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.

1.1. Motivation

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.

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.

2. Decision Making Process

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 and sort that information in a way that fits our purpose.

2.1. Data Processing

With that in mind, using 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 1.

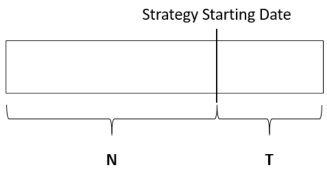

Figure 1: Price data division

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.

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 work.

The first two, 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 2.3). 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.

2.2. Selection based on stocks' relation with the market (CAPM)

The Capital Asset Pricing Model (CAPM) is one of the most widely accepted theoretical assumptions for modelling stock returns [6, 7, 8]. 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

$$r_i = \alpha_i + \beta_i r_m \tag{2}$$

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

$$r_i - r_f = \alpha_i - r_f + \beta_i r_f + \beta_i r_m - \beta_i r_f \quad (3)$$

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

$$r_i = r_f + \beta_i(r_m - r_f) \quad (4)$$

Equation (4) is known in the world of finance as Security Market Line (SML). For investment purposes, what should interest us is the relative variation of r_i and r_m which is given by

$$\frac{\partial r_i}{\partial r_m} = \beta_i \quad (5)$$

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.

2.2.1 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 2.2, 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

$$r_d = \frac{p_d - p_{d-1}}{p_{d-1}} \quad (6)$$

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

$$r_i = \beta_i r_m + \alpha \quad (7)$$

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

$$\beta_i = \frac{COV(r_m, r_i)}{VAR(r_m)} \quad (8)$$

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

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.

2.3. Selection based on stocks' Hurst Exponent value

The Hurst exponent (H) is a statistical test used to measure the long-term memory of a time series. Since its introduction by [9] in 1951, it has been widely used in the field of finance. In 2017, [10], 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.

2.3.1 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 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 low 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.

2.4. 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.

2.4.1 Spread computation

In pursuance of constructing the spread, a linear regression between the prices of the two assets, throughout the training period, can be computed. The resulting weight coefficient of a given pair of stocks (X/Y), B, can be calculated as

$$B = \frac{COV(P_{(X)}, P_{(Y)})}{VAR(P_{(X)})} \tag{9}$$

where $P_{(X)}$ and $P_{(Y)}$ denote the price series for both stocks. In practice, this value can be interpreted as a valuation ratio between the two assets. In other words, in average, each time the X stock values by one dollar, Y stock is expected to value by B dollars.

Having set B, it is now possible to form the spread series of the pair, V_P, as

$$V_P = P_{(Y)} - BP_{(X)} \tag{10}$$

2.5. Filtering Pairs

Cointegration between two time-series can be evaluated according to the definition presented on Equation (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 [11] and the Schmidt–Phillips Test [12]. Through this dissertation's scope, we will only consider the Augmented Dickey-Fuller test (ADF) [13].

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, then 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 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.

2.6. Applying an investment strategy to the filtered pairs

2.6.1 Mean reversion and Z-Score

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

$$\mu_t = \frac{\sum_{k=0}^{14} V_{P(t-k)}}{15} \tag{11}$$

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

$$\sigma_t = \overline{\sigma}_{daily}[t-14;t] \tag{12}$$

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

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

$$z_t = \frac{V_{P(t)} - \mu_t}{\sigma_t} \tag{13}$$

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.

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. Every time the Z-Score is above the threshold value a short position is assumed on the pair's spread. If the Z-Score turns negative, this position is unwind. The inverse methodology is adopted when the Z-Score is below the lower threshold.

Having defined the decision process, it is now time to focus our attention on monitoring the strategies, and quantifying their performance.

3. Results

In this chapter, we will present and analyze, the results of this work. The probabilty of cointegration, the absolute returns and the annualized Sharpe ratios, of all simulations, shall be discussed and compared.

3.1. Cointegrability Results

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 2, shown below, depicts the average percentage of cointegratable pairs, when a N day training period is considered, for each of the three stock choosing methods.

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 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.

3.2. Absolute Returns per Period

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 2, 3 and 4.

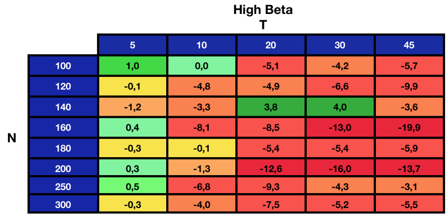

Figure 2: Heat map of the average absolute return, per T day session, in percentage, for pairs formed by combining stocks selected by the high Beta criterion. Row labels in days (N), and the column labels in days (T).

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 3: 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 days (N), and the column labels in days (T).

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 3 than there are on Figure 2. Furthermore, the number of strategies whose losses, in average, surpassed 7.5%, is much higher on Figure 2's case. Highlighted in white (on Figure 3), 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 4: 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 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, annualized Sharpe Ratio, for each strategy.

3.3. 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, 6 and 7, depict the results obtained for the three considered stock selection procedures. Figure 5, shown below, concerns the high Beta procedure.

High Beta

N \ T	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: 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 days (N), and the column labels in days (T).

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 6: 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).

Figure 6, 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 7: 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).

Finally, on Figure 7, 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 $T = 30$).

4. 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 varying parameters, were studied and examined, 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. 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. 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.

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.

5. 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 work' 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.

References

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

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

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

[4] Robert F Engle and Clive WJ Granger. Co-integration and error correction: representation, estimation, and testing. *Econometrica: journal of the Econometric Society*, pages 251–276, 1987.

[5] David A Dickey and Wayne 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.

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

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

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

[9] Harold Edwin Hurst. Long term storage capacity of reservoirs. *Trans. Amer. Soc. Civil Eng*, 116:770–799, 1951.

[10] José Pedro Ramos-Requena, JE Trinidad-Segovia, and MA Sánchez-Granero. Introducing hurst exponent in pair trading. *Physica A: statistical mechanics and its applications*, 488:39–45, 2017.

[11] Graham Elliott, Thomas J Rothenberg, and James H Stock. Efficient tests for an autoregressive unit root. Technical report, National Bureau of Economic Research, 1992.

[12] Denis Kwiatkowski, Peter CB Phillips, Peter Schmidt, and Yongcheol 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.

[13] Peter CB Phillips and Pierre Perron. Testing for a unit root in time series regression. *Biometrika*, 75(2):335–346, 1988.