1 2 9 0

UNIVERSIDADE D COIMBRA

Inês Pinão Fernandes

PAIR TRADING
STRATEGIES AND PERFORMANCE

Dissertação no âmbito do Mestrado em Métodos Quantitativos em Finanças, orientada pelo Professor Doutor Helder Sebastião e pelo Professor Doutor Rui Pascoal e apresentada ao Departamento de Matemática da Faculdadede Ciências e Tecnologia e à Faculdade de Economia.

Março de 2024

Pair Trading: Strategies and Performance

Inês Pinão Fernandes

Master in Quantitative Methods in Finance

Mestrado em Métodos Quantitativos em Finanças

MSc Dissertation | Dissertação de Mestrado

Março 2024

Acknowledgements

I would like to express my deepest gratitude to the following individuals who have played a significant role in the completion of my master's dissertation:

I am sincerely thankful to my thesis supervisors, Professor Helder Sebastião and Professor Rui Pascoal, for their guidance and valuable insights throughout the entire research process.

My heartfelt thanks go to my parents and my sister for their unwavering support, encouragement, and understanding throughout this challenging journey.

To my boyfriend, I am profoundly grateful for your support, love, and understanding throughout the journey of my master's thesis.

I am grateful to my friends, their presence has made this academic pursuit not just successful but also memorable and enriched with shared moments.

Lastly, I want to thank the University of Coimbra, which became not just an academic institution but an enabler for experiences that expanded my horizons and equipped me with a broader perspective as I follow my journey forward.

The work presented in this paper was partially carried out in the scope of the RiskBigData project: Complex Risk Management in the Big Data Regime (PTDC/MAT-APL/1286/2021), financed by the Portuguese Foundation for Science and Technology (FCT).

Abstract

Pair trading, a type of statistical arbitrage strategy, remains an interesting topic in the literature, as it has proven multiple times to be profitable. This dissertation delves into the construction and optimization of pair trading strategies, mixing methods already found in the literature with some original contributions. The prime goal of this dissertation is to identify the best methods for selecting pairs and defining thresholds, thus creating an optimal pair trading strategy. For pair selection, where the objective is to find pairs of assets that have co-movement, various methodologies such as cointegration analysis, Generalized Hurst Exponent (GHE), and additional criteria such as the Sharpe ratio and Ornstein-Uhlenbeck (OU) parameters are explored to enhance the effectiveness of pair trading strategies. In order to define opening and closing profitable thresholds, the OU process with entropy proposed by Yoshikawa is used, also a better adjusted method, such as the Z-score method is applied, with the threshold choice being endogenized by a grid search. The originality is in the use of the OU parameters as additional criteria and the implementation of endogenized Z-score thresholds choice.

In the empirical analysis, a data set comprising 125 assets from Portuguese, Spanish, French, and German markets, spanning from 1990 to 2023, is used. Different combinations of strategies are tried, and the findings showed that GHE-based pair selection is often more profitable than cointegration-based. In the threshold definition, Z-score endogenized thresholds prove to yield higher returns than the Yoshikawa method. Additionally, incorporating OU parameters as additional criteria enhances profitability, particularly when choosing pairs with a high volatility. Despite market illiquidity in the data set, pair trading strategies exhibit a tendency to profit even higher in these conditions. The pair trading strategy also outperformes the benchmarks, performing better under market downturns such as the one that occurred during the 2020 pandemic crisis, proving to be a market-neutral strategy. Overall, pair trading proves to be profitable and surpasses the returns of the buy-and-hold strategy and equal-weighted portfolio used as benchmarks. The strategies in analysis also surpasses the returns presented in the literature, showing that the new implementations, such as the endogenization of parameters and the OU parameters as additional criteria, are profit enhancers.

Keywords: Pair trading, Cointegration, Generalized Hurst exponent, Ornstein-Uhlenbeck process, Performance measures.

Resumo

O *pair trading*, um tipo de estratégia de arbitragem estatística, continua a ser um tópico interessante na literatura, uma vez que já demonstrou várias vezes ser lucrativo. A presente dissertação debruça-se sobre a construção e otimização de estratégias de *pair trading*, misturando métodos já encontrados na literatura com algumas contribuições originais. O objetivo principal desta dissertação é identificar os melhores métodos de seleção de pares e de definição de limiares, criando assim uma estratégia de *pair trading* ótima. Para a seleção de pares, onde o objetivo é encontrar pares de ativos que tenham co-movimento, são exploradas várias metodologias como a análise de cointegração, o expoente generalizado de Hurst (GHE), e critérios adicionais como o rácio de Sharpe e os parâmetros do processo de Ornstein-Uhlenbeck (OU) para aumentar a eficácia das estratégias de *pair trading*. Para determinar limiares de abertura e fecho rentáveis, é utilizado o processo OU com entropia proposto por Yoshikawa e também um método melhor ajustado, como o *Z-score*, sendo neste caso os limiares endogenizados por uma pesquisa em grelha. A originalidade reside na utilização dos parâmetros de OU como critérios adicionais e na implementação da escolha endogenizada dos limiares de *Z-score*.

Na análise empírica, é utilizado um conjunto de dados que inclui 125 ativos dos mercados português, espanhol, francês e alemão, entre 1990 e 2023. São experimentadas diferentes combinações de estratégias e os resultados mostram que a formação de pares baseada no GHE é frequentemente mais rentável do que a baseada na cointegração. Na definição de limiares, os parâmetros endogenizados do *Z-score* provam proporcionar retornos mais elevados do que o método de Yoshikawa. Além disso, a incorporação dos parâmetros de OU como critérios adicionais aumenta a rentabilidade, em especial quando se escolhem pares com uma volatilidade elevada. Apesar da falta de liquidez no conjunto de dados, as estratégias de *pair trading* apresentam uma tendência para obter lucros ainda mais elevados nestas condições. A estratégia de *pair trading* também tem um desempenho superior ao dos *benchmarks*, apresentando um melhor desempenho em períodos de recessão do mercado, como o que ocorreu durante a crise pandémica de 2020, provando ser uma estratégia neutra face ao mercado. Em termos gerais, o *pair trading* revela-se rentável e supera os retornos da estratégia *buy-and-hold* e do portfólio de igual ponderação, utilizados como valores de referência. As estratégias analisadas também superam os retornos apresentados na literatura, mostrando que as novas implementações, como a endogenização de parâmetros e os parâmetros de OU como critérios adicionais, são potenciadores de retorno.

Palavras-Chave: *Pair trading*, Cointegração, Expoente generalizado de Hurst , Processo de Ornstein-Uhlenbeck, Medidas de performance.

Table of contents

List of figures

List of tables

Chapter 1

Introduction

Pair trading is a statistical arbitrage method that has its roots on Wall Street in the 1980s. This trading strategy originated in the work of Morgan Stanley's researcher, Nunzio Tartaglia who worked with a team of physicists, mathematicians, and computer scientists. In 1987, Morgan Stanley profited $50 million with this strategy but discarded it afterwards due to the following years' poor performance. Despite it, pair trading is still a highly well-known strategy nowadays (Gatev et al., 2006) [13].

Pair trading aims at exploiting relative price movements between two related assets. The underlying principle involves identifying pairs of assets whose prices exhibit a certain degree of co-movement over time and strategically capitalizing on the temporary divergences from their historical relationship. An essential characteristic that distinguishes pair trading from conventional trading strategies is its inherent market neutrality. Market neutrality refers to the strategy's ability to minimize exposure to overall market movements. In a pair trading strategy, the focus shifts from predicting the absolute direction of the market to forecasting the relative performance between the assets pair.

The basic assumption in pair trading is that while individual assets may experience significant price fluctuations, the relative price remains more stable. The strategy capitalizes on the mean-reverting nature of asset prices, suggesting that periods of divergence are eventually followed by a reversion to the historical mean. Pair trading involves a three-step procedure: pair selection, threshold definition, and implementation and performance analysis.

In the first step, the challenge lies in identifying suitable pairs that exhibit a meaningful degree of co-movement. Common methodologies employed in this stage include principally cointegration analysis and Generalized Hurst Exponent (GHE), but other methodologies such as correlation or distances analyses have also been applied. These techniques aim at verifying if the selected pairs have a long-term relationship that is mean-reverting. Cointegration in pair trading is a statistical concept that identifies pairs of assets with a long-term relationship, indicating they move together over time. It helps traders select pairs where deviations from their historical relationship are likely to revert to the mean. The GHE in pair trading serves as a measure of the long-term memory or persistence of a time series. A higher Hurst exponent suggests a more persistent trend, while a lower one indicates an anti-persistent behavior, with a tendency for mean reversion.

Additional selection criteria can also be employed, such as the Sharpe ratio, which can be used to select the best in-sample pairs in terms of risk-adjusted profit (although in-sample profit does not guarantee out-of-sample profit).

The second stage involves identifying the thresholds at which the pair's position is open and closed. Choosing optimal thresholds is crucial in implementing profitable strategies as they define entry and exit points, directly influencing trading decisions. Well-selected thresholds help capture meaningful price divergences without generating excessive trades, ensuring the strategy's effectiveness. Precision in the threshold definition minimizes false signals, enhancing the risk-reward ratio and overall performance of the trading strategy. Threshold determination may use machine learning techniques, or more commonly, thresholds are identified using Z-scores. The Z-score quantifies the number of standard deviations a data point is from the mean, offering a standardized metric to identify instances of relative mispricing. Optimization of the Z-score thresholds for initiating and closing the strategy, achieved through a systematic grid search, enhanced the strategy's robustness. Another well-known procedure to define thresholds is by modelling the spread between asset prices using the Ornstein-Uhlenbeck (OU) process. Additionally, entropy measures can be used to measure the uncertainty inherent in the model, providing a more complete understanding of the strategy's reliability in dynamic market conditions.

Another possible additional criteria for pair selection that was implemented in this dissertation and that represents an original aspect that obtained good results is a criteria based on the OU parameters of mean-reversion velocity and volatility. This additional criteria generated good returns.

The last stage involves the implementation of the strategy out-of-sample to assess the profit characterization of the strategies. A comparative analysis between the pair trading strategy and benchmarks such as the traditional buy-and-hold approach or the equally weighted portfolio can be used to assess whether the additional effort and complexity of pair trading lead to superior returns. This comparison clarifies the effectiveness and risk-adjusted returns of pair trading, providing insights into its potential as an alternative investment strategy.

In this dissertation, various strategies are implemented in order to assess their returns, allied with other metrics, to find the optimal pair trading strategy. To perform this, a data set was constructed with the daily prices of 125 assets in the Portuguese, Spanish, French, and German markets. The data was collected between 1990 and 2023 and partitioned into three samples: the training set, the validation set, and the test set. They are utilized to select pairs, determine thresholds, and implement out-of-sample, respectively. The software used to perform the calculations and implement the strategies is Matlab.

The structure of the present dissertation is as follows. Chapter 2 describes the most relevant pair trading strategies found in the literature. Chapter 3 brings a more detailed explanation of important concepts underlying pair trading, such as cointegration, GHE, OU process, and entropy, among others. Chapter 4 presents some important mathematical results to be applied throughout the implementation of pair trading strategies. Chapter 5 delves into the construction of the strategies phase by phase, bringing a complete description of every method employed empirically along with the presentation of the results. Chapter 6 summarizes the main conclusions.

Chapter 2

Literature review

Although pair trading was in place by practitioners in the 1980s, only after 2000 academics begun researching on this topic. Since then the literature has grown steadily almost at an exponential rate.

Gatev et al. (2006) [13] is one of the first articles to address this topic with a visible repercussion in the academic community (although the working paper was written in 1998). The authors selected highly correlated pairs, whose normalized prices had the minimum squared distance. The authors used a threshold of two historical standard deviations. When the spread reached this threshold, the strategy implied taking a long and short positions in the underperforming asset and overperforming asset, respectively. The strategy delivered an annual return of 15.72% for 5 best pairs and 17.28% for 20 best pairs. So, the authors concluded that the strategy was statistically and economically profitable.

Andrade et al. (2005) [2] based their work on the article of Gatev et al. (2006)[13] to choose the 20 pairs with the smallest price spread. The pair selection was reviewed semi-annually. The authors found an annual excess returns of 10.18% , concluding that the strategy was as profitable in the Taiwan and USA market. The authors argued that pair trading profit was a compensation for providing liquidity in the market.

Broussard and Vaihekoski (2012) [3] also used a similar strategy to Gatev et al. (2006) [13]. The threshold to initiate the pair position was initially set at two times the standard deviation of the price differences during the formation period. The study analysed the sensitivity of returns to this threshold ranging from 0.5 to three times the standard deviation. The study concludes that the smaller the threshold, the higher the average return, indicating that lower thresholds allow for potential profits from smaller price divergences. However, it is noted that lower thresholds may result in higher trading costs that could offset the generated returns. The authors obtain a annual return of 30.48% considering committed capital and 69.37% considering a fully invested scheme with equal weighting on pairs. As in Andrade et al. (2005) [2] the authors argued that the high returns were due to market illiquidity.

Engelberg et al. (2008) which study the effects of news, information, and liquidity on the pair trading strategy, follow a pair trading method as proposed by Gatev et al. (2006) [13]. The findings are that pairs from the most illiquid tercile outperform those from the most liquid tercile. The author argues that the short-term profits from pair trading are rewards for providing immediate liquidity and that the long-term profits are larger among illiquid stocks. It was also found that pairs involving smaller, less liquid, and more volatile stocks tend to converge faster after initial divergence.

Jacobs and Weber (2015) [17], which also uses a distance method, present the first cross-country study of pairs trading. They found high returns from pairs trading, particularly pronounced in emerging markets and in markets with a large number of eligible pairs. This profitability is linked to factors such as limits to arbitrage and information overload. The study also covers the U.S. market over a 47-year period, where pair trading yields annualized excess returns of at least 12%.

Do and Faff (2010) [10] re-examine the pair trading strategy proposed by Gatev et al. (2006) [13] and their study was motivated by the fact that very few pair trading constrution methods have stood the test of time and independent scrutiny. The authors extend Gatev et al. (2006) [13] study to data up to June 2008, a decade after the publication of the original research of Gatev et al. (2006) [13] , confirming a continuation of the profitability decline to the point where the strategy's excess return is no longer statistically significant. They associate this decrease to a reduction in arbitrage opportunities during recent years, as measured by the increase in the proportion of pairs that diverge but never converge. Notably, the authors found that the pair trading performance was particularly strong during the market downturn of 2007-09.

Caldeira and Moura (2013) [5] discused the use of cointegration to identify stocks for pairs trading strategies, i.e. identifying stocks whose linear combination exhibits a significant predictable component uncorrelated with market movements. The article mentions that cointegration tests are performed on all possible pairs to determine their suitability for the pair trading strategy. The trading signal was based on the Z-score of the price spread, which is a measure of the distance to the long-term mean in units of long-term standard deviation. The authors considered buy (sell) position in the spread if the Z-score is higher than -2.00 (lower than 2.00). to close the long position the Z-score had to be higher than -0.50, while to close the short position the Z-score has to be inferior to 0.75. Additionally, stop-loss constraints are implemented to prevent large losses, for instance the positions are not kept open for more than 50 days. The article presents out-of-sample results showing that the strategy yields an annual return of 16.39% considering a fully invested schema with equal weighting.

Huck and Afawubo (2015) [15] uses the S&P 500 index to determine if cointegration is a superior method, comparing it to the distance method, Gatev et al. (2006), and the stationarity criterion. The stationarity criterion is based on the fact that the ratio between the two stocks needs to have a constant mean and a constant volatility over time, and a deviation of the price ratio from this equilibrium can be interpreted as a trading opportunity. In this article, the authors select pairs whose price ratio exhibits a unit root, using for that a Dickey and Fuller stationarity test. The author states that both the distance method and stationarity criterion returns are much lower compared to cointegration which returns are greater than 16.56% annually and can rise up to 60% annually, also reducing significantly nonconvergence risk.

Rad et al. (2016) [21] perform a robust study on the performance of three different pairs trading strategies, the distance, cointegration, and copula methods, using a data set including the entire US equity market from 1962 to 2014. Copula is a multivariate cumulative distribution function used to model the dependence between random variables, in this case the two stocks. The findings state that distance methods and cointegration have much higher monthy excess returns than the copula method, eventhough showing better performance for its unconverged pairs. Another conclusion is that the three methods perform better under significant market volatility, but the cointegration is better under periods of market downturn.

Ramos-Requena et al. (2017) [22] proposed a new approach to pair trading using the Hurst exponent as a method for pair selection. The study used data from stocks from Dow Jones Index from 2000 to 2015 and compares the performance of the Hurst approach with other strategies, such as the correlation and distance method. The threshold identification is accomplished via the breakout volatility model (BVM) to generate buy/sell signals for each pair. The upper (lower) threshold is equal to the exponential moving average (EMA) plus (minus) twice the rolling standard deviation of the spread. The results showed that the Hurst approach outperforms the other strategies, especially when the portfolio includes ten or more pairs. The Hurst approach was able to reduce losing operations and increase winning ones, resulting in higher average returns. The procedure is market neutral, which means that a profit is made either the asset price increases or a decreases, and performs well during market crashes. For the GHE pair formation the author obtains a minimum annual return of 34.50% for 5 pairs and a maximum annual return of 49.40% for 25 pairs.

Bui and Slepaczuk (2021) [4] also compares the performance of the three methods, Hurst exponent, correlation, and cointegration, in generating trading strategies. The threshold identification is similar to Ramos-Requena et al. (2017) [22], however the study consider the in-sample optimized values of the threshold parameters, namely the window size of the EMA, the multiplier and the window size of rolling standard deviation. The findings suggest that the Hurst exponent method of pair selection is superior to the cointegration method but not as effective as the correlation method. The authors showed that the results are sensitive to the number of pairs selected and the rebalancing period, but less sensitive to the degree of financial leverage. The paper acknowledges some limitations, including the impact of survival bias due to the lack of full historical data and the use of Pearson correlation, which only measures linear relationships. The study, unlike others before, concluded that the market cannot be outperformed unless in recession periods.

The literature has provided other procedures for choosing thresholds. For instance, Elliott et al. (2005) [11] support theoretically a mean-reverting Gaussian Markov chain to model the price spread and suggests the use of trading strategies based on the first-passage times of the Ornstein-Uhlenbeck process. Do et al. (2006) [9] proposed a new strategy called the stochastic residual spread model of pairs trading. It is a parametric model that aims to identify pairs of assets that are mispriced and profit from the short-term mispricing. The model quantifies the state of mispricing using a residual spread function that takes into account exogenous factors. The spread is assumed to follow an Ornstein-Uhlenbeck process, as in Elliot et al. (2005) [11].

Yoshikawa (2017) [32] highlights the mean reversion feature of the Ornstein-Uhlenbeck process and discusses the application of the relative entropy concept in pair trading. The relative entropy is used as a penalty function to account for model uncertainty in pair trading strategies, arising from the misspecification of model assumptions or incorrect estimation of model parameters. The article proposes a more robust approach to pair trading that maximizes profit while minimizing the relative entropy with respect to the reference measure. The results using data on some Japanese stocks, from May 25, 2015, to September 2, 2015, range from 2.8% and 32.1%, when only four of the stocks used experienced a positive return.

Amer and Islam (2023) [1]explores the use of the Yoshikawa (2017) [32] model, using daily data on 64 companies listed on the Pakistan Stock Exchange (PSX) from 2017 to 2019. The results (returns between 0.2% and 25.2%) show that the entropy approach yielded positive and significant returns compared to the buy and hold strategy. The study also finds that the optimal returns are associated with lower values of the parameter λ. This parameter is present in the formula for computing the optimal boundary points in the context of pair trading, and is defined by Yoshikawa (2017) [32] as the level of trust the agent has in the reference measure. The lower the value of λ, the lower the confidence of the agent on the reference measure as a true probability measure.

Yu et al. (2013) [33] implemented a stochastic mean-variance approach, where the price spread is assumed to follow an Ornstein-Uhlenbeck process. The constrained optimization problem aims at maximizing the mean-variance criterion, which involves minimizing the expected squared deviation of the portfolio value from the expected portfolio value.

Some recent articles propose the use of machine learning techniques to enhance the performance of pair trading. For instance, in Sarmento and Horta (2020) [23] two main issues were addressed: finding profitable pairs while constraining the search space and avoiding long decline periods due to prolonged divergent pairs. The results pointed out that machine learning improved the performance of pair trading. The best profit obtained was 17.0%, with a Sharpe ratio of 3.58.

Chapter 3

Concepts

This chapter discusses important concepts and theorems that are of pivotal importance for pair trading.

3.1 Returns and price spreads

Pair trading involves buying the relative underpriced asset at price P_1 and selling short the relative overpriced asset at price P_2 simultaneously. This means buying a long-short portfolio, with positive position in asset 1 and negative position in asset 2. In pair trading it is common to consider the natural logarithm of prices and use a normalizing constant β. Hence the normalized log-spread at time t is given by:

$$S_t = ln(P_{1,t}) - \beta ln(P_{2,t}) = p_{1,t} - \beta p_{2,t}, \tag{3.1}$$

where P denotes price and p log-price. The parameter β may be interpreted as the amount of proceedings resulting from selling short asset 2 for a unitary amount bought of asset 1. Supposing that the log-prices share a long-run relationship such that $p_{1,t} = a + \beta p_{2,t}$, differentiating $\frac{dP_{1,t}}{P_{1,t}} - \beta \frac{dP_{2,t}}{P_{2,t}} = 0$, or equivalently $\frac{1}{P_{1,t}} dP_{1,t} - \frac{\beta}{P_{2,t}} dP_{2,t} = 0$. So, the quantities are $\frac{1}{P_{1,t}}$ for asset 1 and $-\frac{\beta}{P_{2,t}}$ for asset 2. Hence, the amount allocated to asset 1 is the unity, $\frac{1}{P_{1,t}} P_{1,t} = 1$, while the amount obtained from selling short asset 2 is $\frac{\beta}{P_{2,t}} P_{2,t} = \beta$. If the strategy involves selling short asset 1, then, maintaining the normalisation of prices in order to asset 1, β is the amount invested into asset 2.

The trading return is the spread between the assets, S_t, defined in Equation (3.1), when the market position is closed at a null spread. This is seldom the case. Hence returns are computed as in Vidyamurthy (2004) [30] and Caldeira and Moura (2013) [5]. When the strategy is initiated as a long position in the spread, one has $-S_t$ and when the position is closed one has $+S_t$. The return of the short strategy is given by the sum of the closing spread, $-S_t$, and opening spread, $+S_t$. The daily return, is the sum of the spread at present day, t, and previous day, $t-1$. For the long strategy (long in the spread), this is given by:

$$\begin{aligned} r_t &= -S_{t-1} + S_t \\ &= p_{1,t} - p_{1,t-1} - \beta(p_{2,t} - p_{2,t-1}) \\ &= ln\left(\frac{P_{1,t}}{P_{1,t-1}}\right) - \beta ln\left(\frac{P_{2,t}}{P_{2,t-1}}\right). \end{aligned} \tag{3.2}$$

For the short strategy:

$$\begin{aligned} r_t &= S_{t-1} - S_t \\ &= p_{1,t-1} - p_{1,t} - \beta(p_{2,t-1} - p_{2,t}) \\ &= ln\left(\frac{P_{1,t-1}}{P_{1,t}}\right) - \beta ln\left(\frac{P_{2,t-1}}{P_{2,t}}\right). \end{aligned} \tag{3.3}$$

On the days when the strategy is closed or rebalanced, transaction costs are taken into account. The computations consider proportional transaction costs such that $+2ln\left(\frac{1-c}{1+c}\right)$, where $c = 0.5\%$ is the figure most used in the literature (DeMiguel et al. (2009) [7]). With transaction costs, the price of buying at time $t-1$ an asset with price P_{t-1} becomes $P_{t-1}(1+c)$, and the price from selling at time t is $P_t(1-c)$. By the additive property of the logarithm function one obtains a profit given by the log-returns discounted twice by c because there are two transactions (opening and closing the positon).

Although log-returns are additive in time they are not additive in the asset space. Hence r_t need to be transformed into discrete return, R_t. To calculate the strategies return, the discrete return of each asset in each day t is computed as

$$R_{i,t} = \frac{R_{i,t} - R_{i,t-1}}{R_{i,t-1}}. \tag{3.4}$$

The bijective transformation into continuous returns is given by:

$$R_t = e^{r_t} - 1. \tag{3.5}$$

Let i be a pair trading. The daily return, at time t of a portfolio formed by N pair trades is given by:

$$R_t = \sum_{i=1}^{N} w_{i,t} R_{i,t}, \tag{3.6}$$

where $w_{i,t}$ is the weight of pair i at time t in the portfolio, which are assumed to be $1/N$. By using equal weights, the investment strategy is self-funded in the sense that the investor does ot need to allocate any capital to pursue the investment. Broussard and Vaihekoski (2012) [3], assume that capital is always divided between the pairs that are open on day t. The amount resulted from closing an assets pair is invested into new pairs, and if that pair is reopened, the money is reinvested back by redistributing the capital between all the opened pairs on that day.

3.2 Spurious correlation and cointegration

Spurious correlation refers to the misleading appearance of a statistical relationship between two variables when, in reality, there is no real correlation between them (Simon,1954) [27]. The variables may exhibit a high degree of correlation in the short-term, but this correlation just arises from the existence of trends and does not reflect any signficant relationship between the variables. In a spurious regression the estimated coefficients are non-consistent and the Ordinary Least Squares (OLS) residuals are non-stationary (Hamilton, 1994, 592-593) [14]. Often, to solve the non-stationary feature, time series are over-differentiated eliminating important information and leading to spurious results.

Cointegration is a robust statistical concept that helps understanding the complexities of non-stationary economic and financial time series, which share long-term stochastic relationships (Hamilton, 2006, 571-582) [14]. Cointegration acknowledges that a given linear combination of these variables remain stable throughout time, although the variables may exhibit trends and drifts from their initial values. These stable linear combination is called cointegration vector.

Two non-stationary, integrated of order 1 series, $X_{1,t}$ and $X_{2,t}$, are cointegrated if there is a stochastic linear combination:

$$X_{1,t} = a + bX_{2,t} + u_t, \tag{3.7}$$

that is stationary. a and b represent the cointegration parameters and u_t is the stationary error term. In summary, $X_{1,t}$ and $X_{2,t}$ are integrated of order 1, $X_{1,t}, X_{2,t} \sim I(1)$, and their linear combination, $u_t = X_{1,t} - a - bX_{2,t} \sim I(0)$.

By modeling the long-term equilibrium relationship, one focuses on the possible futures movements of the relative price changes arising from mean reversion dynamics, therefore cointegration is an effective tool for pair trading. When a change in this long term equilibrium relationship is noticed, the strategy (buying the undervalued asset and selling short the overvalued asset) is implemented. If the cointegration relationship holds, the relative prices tend to converge to the long-run relationship and the trading position is closed with a profit.

3.3 Generalized Hurst Exponent

Hurst exponent, $H \in [0,1]$, is a statistic that characterizes the persistence or mean-reversion behavior of time series (Hurst, 1951) [16]. $H < 0.5$ indicates mean-reversion, $H > 0.5$ indicates persistence, and $H = 0.5$ suggests a random or diffusive process.

In pair trading, H is applied to each pair to assess the mean-reversion behavior. The Generalized Hurst Exponent (GHE), denoted by $H(q)$ is calculated using the scaling behavior of the statistic

$$K_q(\tau) = \frac{\mathbb{E}(|X(t+\tau) - X(t)|^q)}{\mathbb{E}(|X(t)|^q)}, \tag{3.8}$$

where $X(t)$ is the time series, τ is the period of observation, and q is a parameter that determines the moments. The scaling behavior of the statistic $K_q(\tau)$ is given by the a power-law, defining $H(q)$ as

$$K_q(\tau) \propto \tau^{qH(q)}, \tag{3.9}$$

where $\propto$ represents a direct proportionality. In financial analysis, particularly in the context of pair trading, the H provides valuable insights. The goal of pair trading strategy, which is to profit from price convergence, is enhanced by a lower Hurst exponent, which implies a larger degree of mean reversion (Ramos-Requena et al., 2017) [22]. In this framework, $q = 1$ and $H(1)$ establishes the scaling characteristics of the absolute deviations of prices.

3.4 Ornstein-Uhlenbeck process

The Ornstein-Uhlenbeck (OU) process models a continuous-time random time series $X(t)$ with mean reversion. Thus process plays a fundamental role in understanding the dynamics of interest rates and other assets that tend to return to an equilibrium state (Ornstein and Uhlenbeck, 1930) [29].

The OU process is typically described by the following stochastic differential equation:

$$dX_t = \alpha(\mu - X_t)dt + \sigma dW_t, \tag{3.10}$$

where X_t represents the random variable of interest, α is a positive constant that determines the speed of the mean reversion dynamics, μ is the long-term mean or equilibrium level of X_t, σ is the volatility or randomness factor representing the size of the stochastic fluctuations, and W is a Wiener process or Brownian motion, representing noise.

Sometimes, the OU process is presented as

$$dY_t = -\alpha Y_t dt + \sigma dW_t, \quad Y_0 = 0 \quad , \tag{3.11}$$

where $Y_t = X_t - \mu$.

3.5 Entropy

This concept was first introduced in physics by Rudolf Clausius in the 1850s (Cropper, 1986) [6]. It is a corollary of applying the second law of thermodynamics to each infinitesimal stage of the heat transfer process. The generalized inequality of Clausius is the following:

$$dS \geq \frac{\delta Q}{Temp}. \tag{3.12}$$

In this equation dS denotes an infinitesimal change in the entropy of a system in any process that occurs in a closed system, δQ denotes the heat transferred, and $Temp$ denotes the system's temperature. As such, entropy tends to increase in isolated systems, explaining the irreversibly in natural processes.

Entropy, as defined by Boltzmann and Gibbs (Jaynes (1965) [18]), is a fundamental concept in thermodynamics and statistical mechanics playing a crucial role in understanding the behavior of physical systems. It is closely related to the idea of energy dispersion and represents a measurement of the disorder or unpredictability in a system. In the late 19th century, the physicist Ludwig Boltzmann made an important impact on the progress of statistical mechanics. The author established a statistical connection between a system's macroscopic characteristics and the microscopic behaviors of its component parts, showing that entropy measures the degree of imprecision of a system's microstate. Higher entropy implies more disorder and unpredictability.

Shannon and Weaver (1949) [24] introduced the concept of entropy as a fundamental measure of uncertainty in the information content of a communication system. Here communication means exchange of information between a sender and a receiver. Entropy became a fundamental idea in information theory, used to express how much randomness or uncertainty there is in a random variable or message traveling across a communication channel. The authors showed that the amount of information conveyed in a message depends on the degree of uncertainty it resolves for the receiver. Messages that reduce uncertainty carry less information. Entropy, hereafter denoted as $En(x)$, is used to quantify the uncertainty associated with a random variable x, and it is defined as the average amount of information needed to describe the outcome of x. Mathematically, it can be expressed as:

$$En_X = -\sum[pr(X)\ln(pr(X))], \tag{3.13}$$

$pr(X)$ is the probability of the outcome X.

This leads to the the statement that a random variable with a uniform distribution (where all outcomes are equally likely) has maximum entropy. If the source of information makes a biased choice (some possible outcomes have a larger probability than others), the entropy is lower. Also, an increase in the number of potential outcomes leads to a higher entropy.

Entropy is useful in the context of pair trading because model uncertainty is a major issue. An inaccurate probability creates uncertainty, which may result in a significant loss. Yoshikawa (2017) [32] uses the concept of relative entropy as a fundamental tool to overcome this uncertainty. Relative entropy is viewed as a penalty function to measure the misspecifications of the model. A more informed threshold choice can be made, and risks can be more effectively managed by quantifying this difference. By providing a strong foundation for resolving the problems caused by the uncertainty in the probability measure, relative entropy ultimately improves the performance and profitability of pair trading.

Relative entropy, which was originally proposed by Kullback and Leibler (1951) [19], is defined in Yoshikawa (2017) [32] as

$$En_X[\mathbb{Q}|\mathbb{P}] := \begin{cases} \mathbb{E}_X^{\mathbb{Q}}[ln(\frac{d\mathbb{Q}}{d\mathbb{P}})], & \mathbb{Q} \in \mathcal{Q} \\ +\infty, & otherwise \end{cases} . \tag{3.14}$$

Since this is a measure of the divergence between the two probability measures, by minimizing the relative entropy with respect to the reference measure $\mathbb{P}$, $\mathbb{Q}$ maximizes the expected return of the strategy, because it loosens the confidence degree that the agent has in the reference probability measure.

3.6 Itô's formula

Ito's formula is a key concept in stochastic calculus, allowing to extend the chain rule of deterministic differential equations to stochastic differential equations (SDEs), particularly those involving a Brownian motion, dW_t. The straightforward chain rule for deterministic differential equations is (Oksendal, 2013, 44-45) [20]:

$$d\phi(t,x_t) = \frac{\partial\phi(t,x_t)}{\partial t}dt + \frac{\partial\phi(t,x_t)}{\partial x}dx_t \quad , \tag{3.15}$$

where x_t is a deterministic differential equation solution. Considering X_t the unique solution of a stochastic differential equation, Ito's formula is given by accomplishing that:

$$d\phi(t,X_t) = \frac{\partial \phi(t,X_t)}{\partial t}dt + \frac{\partial \phi(t,X_t)}{\partial x}dX_t + \frac{1}{2}\frac{\partial^2 \phi(t,X_t)}{\partial X^2}d\langle X\rangle_t, \tag{3.16}$$

where $d\langle X\rangle_t = dX_t dX_t$ is the quadratic variation, which by definition (Oksendal, 2013, 56-57) [20], is:

$$\langle X\rangle_t = \lim_{\Delta t_k \to 0} \sum_{t_k \leq t} |X_{t_{k+1}} - X_{t_k}|^2 \tag{3.17}$$

where $0 = t_1 < t_2 ... < t_n = t$ and $\Delta t_k = t_{k+1} - t_k$. The Itô formula for the case in which dX_t is the diffusion process, given by:

$$dX_t = \mu(t,X_t)dt + \sigma(t,X_t)dW_t. \tag{3.18}$$

Therefore

$$d\phi(t,X_t) = \Big(\frac{\partial \phi(t,X_t)}{\partial t} + \mu(t,X_t)\frac{\partial \phi(t,X_t)}{\partial x} + \frac{1}{2}\sigma^2(t,X_t)\frac{\partial^2 \phi(t,X_t)}{\partial X^2}\Big)dt + \sigma\frac{\partial \phi(t,X_t)}{\partial x}dW_t, \tag{3.19}$$

where $\mu(t,X_t)$ is the expected trend of X_t, $\sigma(t,X_t)$ is the volatility, and W_t is the Brownian motion. In Itô's formula, ϕ is a function of X and t, showing how changes in these values affect the overall function. The quadratic variation term in Itô's calculus depends on the second order derivative of ϕ with respect to x. Ito's formula is distinguished from the common chain rule by this term, which can be calculated according to:

$$dtdt = dW_t dt = dtdW_t = 0, \quad dW_t dW_t = dt. \tag{3.20}$$

This concept is key to the derivation of the Yoshikawa (2017) [32] method, which is presented in Appendix A.

3.7 Girsanov theorem

Girsanov theorem (Oksendal, 2013, 153-160) [20] addresses changes in SDEs resulting from a change in the probability measure. This theorem is based on the observation that the SDE drift depends on the probability measure $\mathbb{P}$ in the probability space $(\Omega, \mathscr{F}_t, \mathbb{P})$, but the diffusion coefficient does not. Thus, the Girsanov theorem can be used to change an SDE's drift coefficient. Two probability measures, $\mathbb{P}$ and $\mathbb{Q}$, in the probability spaces $(\Omega, \mathscr{F}_t, \mathbb{P})$ and $(\Omega, \mathscr{F}_t, \mathbb{Q})$ are equivalent if and only if $\mathbb{P}(A) = 0 \Leftrightarrow \mathbb{Q}(A) = 0$. In the presence of two equivalent processes, there is a process M_t, called the Radon-Nikodym derivative of $\mathbb{Q}$ with respect to $\mathbb{P}$, restricted to the filtration $\mathscr{F}_t$, which is a martingale on $(\Omega, \mathscr{F}_t, \mathbb{P})$ such that:

$$\mathbb{Q}(A) = \int_A M_t(\omega)d\mathbb{P}(\omega), A \in \mathscr{F}_t. \tag{3.21}$$

The process M_t is therefore denoted as:

$$M_t = \frac{d\mathbb{Q}}{d\mathbb{P}}\bigg|_{\mathscr{F}_t}. \tag{3.22}$$

When changing from $\mathbb{P}$ to $\mathbb{Q}$, the expected value of a random variable X, verifies that

$$E_{\mathbb{Q}}[X] = E_{\mathbb{P}}\left[X\frac{d\mathbb{Q}}{d\mathbb{P}}\right]. \tag{3.23}$$

The Girsanov theorem enunciates that a SDE under $\mathbb{P}$, with drift $f(t,\omega)$

$$dX_t(\omega) = f(t,\omega)dt + \sigma(t,\omega)dW_t(\omega), \tag{3.24}$$

where W is a Brownian motion, may be converted into a new SDE under $\mathbb{Q}$ with a new drift $q(x,\omega)$ as:

$$dX_t(\omega) = q(t,\omega)dt + \sigma(t,\omega)dW_t{}^*(\omega), \tag{3.25}$$

where W^* is a Brownian motion under $\mathbb{Q}$. $\mathbb{Q}$ is such that

$$M_t = exp\left(-\int_0^t u(s,\omega)dW_s - \frac{1}{2}\int_0^t u^2(s,\omega)ds\right), \quad u(t,\omega) = \frac{f(t,\omega)-q(t,\omega)}{\sigma(t,\omega)}. \tag{3.26}$$

Where the Brownian motion under $\mathbb{Q}$ verifies:

$$W^*{}_t(\omega) = -\int_0^t u(s,\omega)ds + W_t(\omega). \tag{3.27}$$

This theorem is of particular interest in finance, allowing to move from the real-world to a risk-neural framework. Like Itô's formula, Girsanov theorem is of particular importance in the derivation of the Yoshikawa (2017) [32] method.

3.8 Benchmark investment strategies

In this section, the two benchmarks used for evaluating the performance of the pair trading strategy are the Buy-and-Hold strategy ($B\&H$) and the equal-weighted portfolio, $1/N$ (where N is the total number of assets in the portfolio).

The $B\&H$ is a passive long-term investment strategy. It involves diversified portfolio and holding it for an extended period, often for several years or even decades. This strategy is based on the belief that over time, the value of the portfolio will appreciate due to the risk premium despite short-term market fluctuations.

This strategy is best suited for long-term horizons, i.e. these investors are generally not concerned with short-term market volatility. When using $B\&H$, the investor should have a high tolerance for market fluctuations, as the portfolios may go through periods of decline. What matters is the long-term return. This strategy is often used as a benchmark for assessing other investment strategies. This comparison helps to understand whether a more active or tactical approach outperforms or underperforms the naive strategy. For instance, active trading may yield higher returns in some years, but this comes with higher trading costs and greater risk due to frequent trading. Comparisons often include trading costs and risk measurements to evaluate the overall cost-effectiveness of different approaches compared to the $B\&H$ strategy.

The equal-weighted portfolio is an investment strategy where the assets in the portfolio have equal weights, i.e. the same amount of capital is allocated to every asset regardless of its particular features or market value. An equally weighted portfolio makes sense since it is straightforward strategy and tends to offer balanced exposure across all of the assets in the portfolio. The portfolio seeks to mitigate the investment risk associated with heavily weighting a small number of particular assets by allocating equal weights to each asset. The ability of an evenly weighted portfolio to beat cap-weighted benchmarks is one of its main advantages (DeMiguel et al., 2009) [7], especially when smaller or less liquid assets outperform larger ones. $1/N$ portfolios do, however have the disadvantage of allocating the same amount of money to every asset, regardless of its risk or market dynamics,hence they tend to be more volatile than cap-weighted portfolios.

The returns calculations for both strategies are very similar, with the main differences residing in the weights calculations and the transaction costs. These strategies are applied to the 125 assets in the data set during the test period, the same period where the pair trading strategies are implemented.

In the case of the $1/N$ portfolio, when prices change there is a need to rebalancing the portfolio, hence transaction costs are considered. In the event that one of the assets goes out of the database (e.g., bunkrupcy), the weights are recomputed. In this case, a term $-2c$ is added to $R_{i,t}$, Equation (3.4), to account for the transaction costs of opening and closing the position in the asset. As for the pair trading $c = 0.5\%$.

The daily average return of the $1/N$ portfolio is given by

$$R_t = \frac{1}{N}\sum_{i=1}^{N} R_{i,t}, \tag{3.28}$$

which can be 125 or less in the case where an asset is bankrupt during the test period. The *B&H* strategy has a different weighting scheme given by:

$$w_i = \frac{C_i}{\sum_{j=1}^{N} C_j}, \tag{3.29}$$

where C_i is the market capitalization of asset i. The market capitalizaton of each asset was collected from the Eikon data base in the first day of the test sample, i.e., May 1, 2015. Weights computed from Equation (3.29) are then used in Equation (3.6).

The comparison between these benchmarks and the pair trading strategies are in Subsection 5.4.

Chapter 4

Theoretical approach

One of the methods implemented in this study for the threshold definition is the one proposed by Yoshikawa (2017) [32], which uses an Ornstein-Uhlenbeck (OU) process to model the spread between two assets, adding entropy as a penalty function to account for the model uncertainty. This chapter highlights the main features of this model, which proposes a way to obtain the optimal boundaries for the strategies aiming at their profitability. The Yoshikawa (2017) [32] procedure implies a first step which consists at estimating the OU process parameters. The parameters of the OU process; the mean reversion velocity, α, and the volatility, σ, can be a good additional criteria to select pairs that supposedly lead to good results. This issue is further elaborated in Chapter 5. In Section 4.2, the estimators for the parameters of the OU are presented.

4.1 Yoshikawa method

Yoshikawa (2017) [32] proposed the use OU processes to model the spread between two assets. That article presents a formulation that is useful to calculate the optimal threshold of the strategy, b^*, which in a first stage, is the output of the following optimal stopping problem:

$$v^0(x) = \sup_{\tau \in T} \quad \mathbb{E}_x[e^{-\rho\tau} Y_\tau], \tag{4.1}$$

where T is the set of all possible stopping times of a Wiener process, W, ρ is the discount rate, and Y_t is an OU process defined in Equation (3.11). It is logically assumed that $\mathbb{E}_\tau < \infty$ for all $\tau \in T$. This problem has the discounted return as the objective function, and the spread at opening time is equal to the return if the closing happens when the spread is zero. The strategy proposed is to start a short position in the pair if the spread is higher than b^* closing the position when the spread reverts to its mean. The long position is started when the spread is lower than its mean and closed when the spread is higher then b^*. This optimization problem has the following solution:

$$\frac{\sigma}{\sqrt{2\alpha}} \frac{\int_0^\infty u^{\frac{\rho}{\alpha}-1} e^{\frac{\sqrt{2\alpha}}{\sigma} ub^* - \frac{u^2}{2}} du}{\int_0^\infty u^{\frac{\rho}{\alpha}} e^{\frac{\sqrt{2\alpha}}{\sigma} ub^* - \frac{u^2}{2}} du} = b^*. \tag{4.2}$$

This is the formula for the optimal threshold to initiate the pair trading position. Yoshikawa (2017) [32] considers that this value is uncertain because there is uncertainty about the probability measure

$\mathbb{P}$. Entropy can be used as a penalty function to account for the model uncertainty in the threshold calculations. This can be done by solving the following optimal problem:

$$\underset{\tau \in T}{sup} \underset{\mathbb{Q} \in \mathscr{Q}}{inf} \quad \mathbb{E}_x^{\mathbb{Q}}[e^{-\rho(\tau-t)}Y_\tau], \tag{4.3}$$

where $\mathscr{Q}$ is a class of probability measures on $(\Omega, \mathscr{F})$. This problem is called the maxmin expected utility, but as it most often implies that the investor should not enter the market. Therefore in this dissertation a more flexible approach is considered, conditioning the problem with $x = Y_t$ at time t. Obtaining:

$$v(t,x) = \underset{\tau \in T}{sup} \quad \mathbb{E}_x^{\mathbb{Q}}[e^{-\rho(\tau-t)}Y_\tau], \tag{4.4}$$

where $\mathbb{Q}$ is the solution of

$$\underset{\mathbb{Q} \in \mathscr{Q}}{inf} \{\mathbb{E}_x^{\mathbb{Q}}[e^{-\rho(\tau-t)}Y_\tau] + \lambda e^{-\rho(\tau-t)} En_x[\mathbb{Q}|\mathbb{P}]\}. \tag{4.5}$$

In the previous equation λ is a positive constant that reflects the confidence of the investor in the measure $\mathbb{P}$. When $\lambda \to \infty$ the investor has full confidence in the measure, while when $\lambda \to 0$ the investor has no confidence in the measure. $En[\mathbb{Q}|\mathbb{P}]$ is the relative entropy between the measures $\mathbb{Q}$ and $\mathbb{P}$, defined by Equation (3.14). The solution to this problem is:

$$ln(b_t) + \frac{\alpha}{\sigma^2}\frac{\rho}{\alpha-\rho}(g_t - b_t)^2 = ln(b^*) + \frac{\alpha}{\sigma^2}\frac{\rho}{\alpha-\rho}(b^*)^2, \tag{4.6}$$

where $g_t = -\frac{\sigma^2}{\lambda} t e^{-\alpha t}$. From this equation one can obtain the threshold accounting with model uncertainty b_t. This threshold allows the definition of a strategy in the same way as the one for b^*. The threshold b^* is equivalent to the case $\lambda \to \infty$, where the trust in measure $\mathbb{P}$ is total, and therefore $g_t \to 0$ and $t \to 0$, $b_0 = b^*$. For a detailed deviation of the solutions see Appendix A. It is important to recall the correction to Theorem 1 of Yoshikawa (2017) [32] in Equation (4.6), also better explained in Appendix A.

4.2 Ornstein-Uhlenbeck parameters estimation

To estimate the OU parameters, μ, α, σ, the maximum likelihood method (MLE) is implemented (Suwarman, 2019) [28]. This is a two-step procedure. First, one must compute the likelihood function, and then the values that maximize this function.

In the OU process, whose expression is presented in Equation (3.10), the process X_t is the spread as defined in Equation (3.1). The OU process assumes a Gaussian distribution, so first the likelihood function is computed for the Gaussian density function and only then applied to the OU process. The density function of a Gaussian distribution is given by:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2}, \quad X \sim N(\mu, \sigma^2), \tag{4.7}$$

where μ and σ^2 are the mean and variance, respectively. The likelihood function is given by:

$$L(x_1,...,x_n;\theta) = f(x_1;\theta)\cdot ... \cdot f(x_n;\theta) = \prod_{i=1}^{n} f(x_1;\theta), \tag{4.8}$$

where $x = \{x_1,...,x_n\}$ is the data sample with length n, where all the variables are independent and identically distributed (i.i.d), $\theta = \{\mu,\sigma^2\}$.

To estimate the parameter θ, the likelihood function must be maximized. This is performed analytically, but if not possible, a numerical approach must be used. In order to simplify the computations, the log-likelihood function is used. The log-likelihood function is given by:

$$l(x_1,...,x_n;\theta) = log(L(x_1,...,x_n;\theta)) = log\Big(\prod_{i=1}^{n} f(x_1;\theta)\Big) = \sum_{i=1}^{n} log(f(x_1;\theta)), \tag{4.9}$$

which can be rewritten for the Gaussian distribution as

$$l(x_1,...,x_n;\mu,\sigma^2) = -\frac{n}{2}log(\sigma^2) - \frac{n}{2}log(2\pi) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2. \tag{4.10}$$

As mentioned before the OU process follows a Gaussian distribution, which it follows from its integral form:

$$X_t = X_0e^{-\alpha t} + \mu(1-e^{-\alpha t}) + \sigma e^{-\alpha t}\int_0^t e^{\alpha s}dW_s. \tag{4.11}$$

The process $X_t \sim N(m_t, Var_t)$, with the parameter defined analytically as:

$$m_t = \mathbb{E}[X_t] = X_0e^{-\alpha t} + \mu(1-e^{-\alpha t}), \tag{4.12}$$

and

$$Var_t = \mathbb{E}[X_t^2] - E[X_t]^2 = \frac{\sigma^2}{2\alpha}(1-e^{-2\alpha t}). \tag{4.13}$$

For $t = t_i$ these expressions are rewritten as:

$$m_{t_i} = \mathbb{E}[X_{t_i}] = X_{t_{i-1}}e^{-\alpha t_i} + \mu(1-e^{-\alpha t_i}), \tag{4.14}$$

and

$$Var_{t_i} = \mathbb{E}[X_{t_i}^2] - \mathbb{E}[X_{t_i}]^2 = \frac{\sigma^2}{2\alpha}(1-e^{-2\alpha t_i}). \tag{4.15}$$

where $\mathbf{X} = \{X_{t_0},...,X_{t_n}\}$ is the vector of the spreads with n values that are equidistant by $\Delta t = t_i - t_{i-1}$ in the time interval $t = \{t_1,...,t_n\}$.

Applying this expressions on the log-likelihood for the Gaussian distribution, the formula for the OU process is given by

$$\begin{aligned} l(x_1,...,x_n;\alpha,\mu,\sigma^2) = &-nln\Big(\frac{\sigma^2}{2\alpha}(1-e^{-2\alpha\Delta t})\Big) - \frac{n}{2}ln(2\pi) \\ &- \frac{2\alpha}{2\sigma^2(1-e^{-2\alpha\Delta t})}\sum_{i=2}^{n}\Big(X_i - X_{i-1}e^{-\alpha\Delta t} - \mu(1-e^{-\alpha\Delta t})\Big)^2. \end{aligned} \tag{4.16}$$

Deriving this function and by solving the following equation system:

$$\begin{cases} \frac{\partial}{\partial \alpha} l(x_1,...,x_n;\alpha,\mu,\sigma^2) = 0 \\ \frac{\partial}{\partial \mu} l(x_1,...,x_n;\alpha,\mu,\sigma^2) = 0 \\ \frac{\partial}{\partial \sigma^2} l(x_1,...,x_n;\alpha,\mu,\sigma^2) = 0 \end{cases},$$

one obtains the estimator of each parameter:

$$\hat{\alpha} = -\frac{1}{\Delta t} ln\Big(\frac{\sum_{i=2}^n X_{i-1}X_i - \mu\sum_{i=2}^n X_{i-1} - \mu\sum_{i=2}^n X_i + n\mu^2}{\sum_{i=2}^n X_{i-1}^2 - 2\mu\sum_{i=2}^n X_{i-1} + n\mu^2}\Big), \tag{4.17}$$

$$\hat{\mu} = \frac{\sum_{i=2}^n X_i \sum_{i=2}^n X_{i-1}^2 - \sum_{i=2}^n X_{i-1}\sum_{i=2}^n X_{i-1}X_i}{n\sum_{i=2}^n X_{i-1}^2\sum_{i=2}^n X_i^2 - ((\sum_{i=2}^n X_{i-1})^2 - \sum_{i=2}^n X_{i-1}\sum_{i=2}^n X_i)}, \tag{4.18}$$

$$\hat{\sigma}^2 = \frac{1}{n}\Big(\sum_{i=2}^n X_i^2 - 2\xi\sum_{i=2}^n X_{i-1}X_i + \xi^2\sum_{i=2}^n X_{i-1}^2 - 2\mu(1-\xi)(\sum_{i=2}^n X_i - \xi\sum_{i=2}^n X_{i-1}) + n\mu^2(1-\xi)^2\Big)\frac{2\alpha}{1-\xi}, \tag{4.19}$$

where $\xi = e^{-\alpha\Delta t}$.

Chapter 5

Empirical analysis

This dissertation considers several pair trading strategies. In some strategies additional selection criteria are added. The pair selection is done using cointegration and the generalized Hurst exponent (GHE). The additional criteria are the Sharpe ratio and the mean reversion velocity parameter, α, combined with the volatility parameter, σ, from the Ornstein-Uhlenbeck (OU) process. The threshold choice is also made in different ways, such as the Z-score and the entropy OU method proposed by Yoshikawa (2017) [32]. In this section, the methods employed in every phase of these strategies are explained in detail, along with a description of the used data set. Table 5.1 presents the strategies to be implemented.

Table 5.1 Strategies description

Strategy	First layer pair selection	Second layer pair selection	Threshold definition
C-Z	Cointegration	–	Z-score
C-SR-Z	Cointegration	Sharpe Ratio	Z-score
C-OU-Z	Cointegration	Parameters of the OU process	Z-score
C-OU-Y	Cointegration	Parameters of the OU process	Yoshikawa method
C-Y	Cointegration	–	Yoshikawa method
H-Z	GHE	–	Z-score
H-SR-Z	GHE	Sharpe Ratio	Z-score
H-OU-Z	GHE	Parameters of the OU process	Z-score
H-OU-Y	GHE	Parameters of the OU process	Yoshikawa method
H-Y	GHE	–	Yoshikawa method

Notes: This table presents the ten pair trading strategies methods used in the empirical study. It identifies the pair selection criteria and the thresholds determination. The first column shows, using acronyms, the strategies. Those with just two letters use only a first layer for pair selection. GHE is the acronym for generalized Hurst exponent and OU is the acronym for Ornstein-Uhlenbeck process.

Throughout this dissertation, one endogenized some of the parameters of these strategies. The method of endogenizing a value consists of taking that parameter as a variable and performing a grid search on pre-defined values of the parameter in order to assess the optimal value.

5.1 Data description

The data used in this study consists on the daily total return index of the 125 most liquid Portuguese, Spanish, French and German stocks, retrieved from the Eikon database. The total return index is constructed considering the closing price adjusted to management events, (e.g. stock splits) and dividends assuming their reinvestment into the stock. For simplicity purposes hereafter the total return index is referred to as price and denoted by P, while its natural logarithm is denoted by p. The sample covers the period from January 1, 1990, to October 9, 2023. The firms were selected in the first day of the sample, hence there is no survival bias. German and French firms are the constituent stocks of the DAX 30 and CAC 40 indexes, respectively, while the Portuguese and Spanish firms are the 20 and 35 firms, respectively, with the higher market capitalization in those national stock markets on the first day of the sample (the national benchmark indexes, PSI 20 and IBEX 35, were only launched in 1992). The list of stocks is presented in Appendix B.

The sample was divided into three sub-samples: training (one-half of the observations, from January 1, 1990, to November 20, 2006), used to pair selection, validation (November 21, 2006. to April 30, 2015) used for define the best thresholds for each strategy, and test (May 1, 2015, to October 9, 2023) used to implement and assess the trading strategies out-of-sample. Figure 5.1 illustrates the data partition.

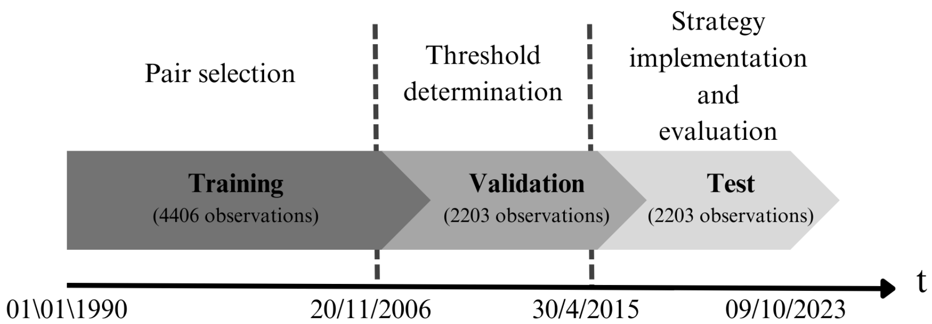

Figure 5.1 Sample partition

In the event of a merger, acquisition, replacement, or even bankruptcy in the training or validation sub-samples, the stock was replaced by the most liquid company in the market in the beginning of the sample not yet in the database. In the case of a firm no longer available at some point in the test subsample, the firm remains in the empirical study until that point in time, guaranteeing the mitigation of survival bias. This event happened for ten Portuguese and three Spanish firms, while there are no cases in the French and German samples.

In this dissertation, when referring to the in-sample, the data sets considered are the training set and validation set, meaning that the data used in the implementation is not included, preventing data mining. Out-of-sample refers to the test data set, only used to implement the strategy.

5.2 Pair selection

5.2.1 First layer criteria

The two methods used for the selection of pairs, cointegration and the GHE, are presented in this section.

Cointegration method

The pairs selected by this method are the ones that show significant cointegration, which is assessed using the Engle-Granger two-step methodology applied to the log-prices. Non-spurious cointegration means that the series are integrated of order one, $I(1)$, and share the same stochastic trend (Wooldridge, 2012, p. 644-648) [31]. Firstly, the Augmented Dickey-Fuller test (ADF) is applied to each time series. This test is suitable for complex processes, such as autoregressive moving average (ARMA) processes with undetermined orders (Zivot and Wang, 2006, p. 120-121) [34]. In this study, the ADF tests the null hypothesis that p_t is $I(k), k \geq 1$, meaning it has a unit root, against the alternative that it is integrated of order 0 ($I(0)$), indicating stationarity. The ADF(p) uses the following regression:

$$p_t = \alpha + \Phi y_{t-1} + \sum_{j=1}^{p} \psi_j \Delta p_{t-j} + \varepsilon_t, \tag{5.1}$$

where α is a constant, Φ is the first order autoregressive coefficient, Δy_{t-j} captures the ARMA structure of the errors, with a lag length p chosen to ensure serially uncorrelated errors, and ε_t is the error term, assumed to be homoscedastic. The number of lags is chosen according to the Akaike information criterion (AIC). Under the null hypothesis that y_t is $I(1)$, Φ is not statistically different from the unity. The ADF test relies on t-statistics

$$ADF = \frac{\hat{\Phi} - 1}{\hat{se}(\Phi)}, \tag{5.2}$$

where $\hat{se}(\Phi)$ is the estimated standard error. It is considered that the series are not integrated if the p-value is lower than 5%.

The ADF(p) suggested that all log-prices were $I(1)$, hence the series are differentiated, $r_t = p_t - p_{t-1}$. This new time series correspond to the continuously compounded returns. The ADF(p) rejected the null hypothesis, suggesting that the continuous returns are stationary.

Secondly, the Engle-Granger test is applied to the initial $I(1)$ series. The test uses a simple regression between the two log-price series estimated by OLS:

$$p_{1,t} = a + bp_{2,t} + u_t, \quad p_1, p_2 \sim I(1). \tag{5.3}$$

The test assesses the stationarity of residuals $\hat{u}_t$ using the ADF test. The null hypothesis is that the variables are not cointegrated.

The Engle-Granger methodology was applied to all possible pairs of the 125 stocks, i.e. to 7,750 pairs. As a rule of thumb it is considered that pairs are cointegrated if the p-value of the Engle-Granger test is lower than 1%.

Figure 5.2 highlights the pairs for which the p-value of the Engle-Granger test is lower than 1%. It was possible to find 411 cointegrated pairs, i.e. 5.3% of all pairs.

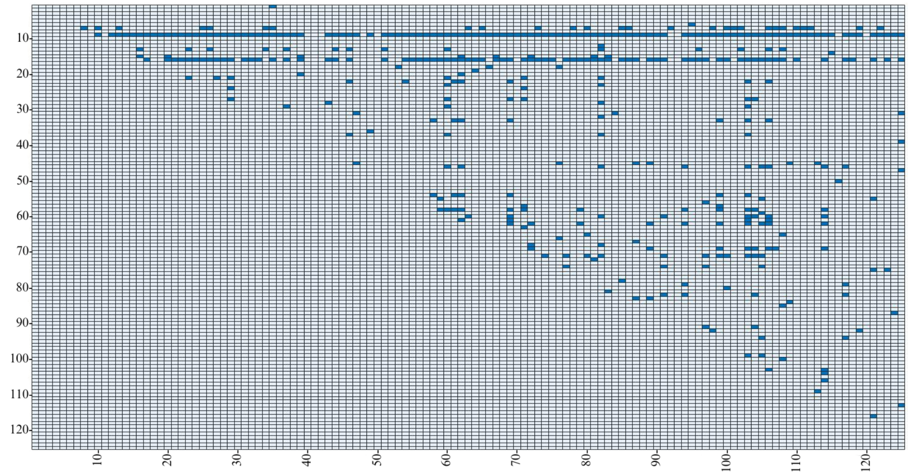

Figure 5.2 Cointegrated pairs (blue dots) with a p-value of the Engle-Granger test lower than 1%

Generalized Hurst exponent method (GHE)

The GHE is a valuable tool to identify pairs with mean reversion properties, which are carachterized by a low GHE.

The application of GHE for pair trading begins by defining the spread between log-prices as in Equation (3.1). Ramos-Requena et al. (2017) [22] proposes

$$\beta = \frac{std(r_{1,t})}{std(r_{2,t})}, \tag{5.4}$$

whith $t \in [0, N_T]$, being N_T the number of days in the sample, and $std(r_{i,t})$ the standard deviation of the log-retun at time t. The parameter β guarantees that the volatilities of the positions in both stocks are equal and provides a fast and efficient way to normalize the stocks in each pair.

The pair selection involves calculating the Hurst exponent for each potential pair using the price spread and select the ones with an Hurst exponent below a given threshold equal or lower than 0.45. This upper limit was found by a grid search on the interval $[0.05, 0.5]$ with a step size of 0.05, and computing the overall return of the pairs that meet the threshold. The return was calculated as in Section 3.1, and the data used in the calculation was the training set. The value that reached the highest return in-sample was 0.45. Only 792 pairs (10.22% of the total number of pairs) with a Hurst exponent inferior to this value were selected. Figure 5.3 highlights the pairs with a GHE lower than 0.45.

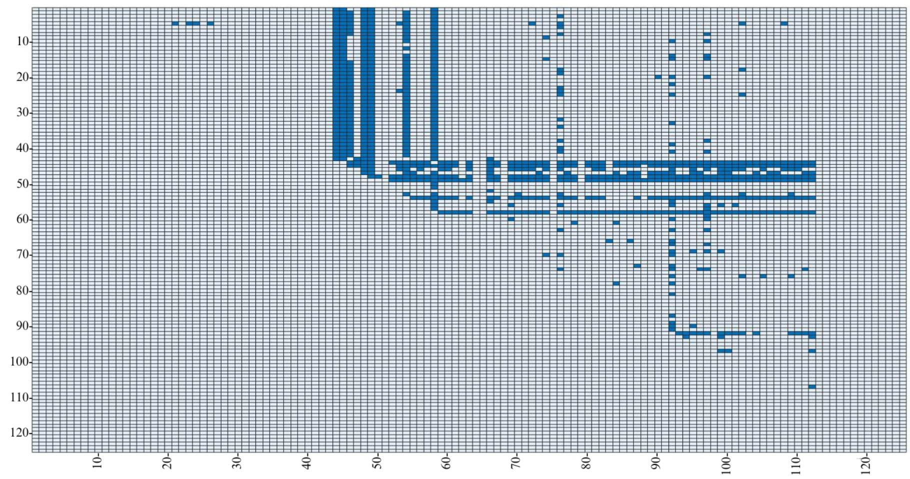

Figure 5.3 GHE pairs (blue dots) with a Hurst exponent lower than 0.45

5.2.2 Second layer criteria

As the number of pairs selected by the two previous methods is quite high, additional selection criteria is usually used to improve the pair selection process. This study uses as additional criteria the Sharpe ratio, *SR* (see Caldeira and Moura [5]) and the mean reversion velocity (α) and volatility (σ) of the OU process, introduced in Section 3.4.

Sharpe ratio

The Sharpe ratio (Sharpe, 1998) [25], is a widely used metric in finance to assess the risk-adjusted performance of an investment or trading strategy. It quantifies the excess return earned per unit of risk, providing a valuable measure for comparing the performance of different strategies or portfolios. Rational investors seek strategies with higher Sharpe ratios, as they suggest better returns for the level of risk taken.

The Sharpe ratio is defined as:

$$SR_i = \frac{\overline{R}_i - R_f}{\sigma_i}, \tag{5.5}$$

where $\overline{R}_i$ is the average daily return of the portfolio or strategy i, R_f is the risk free rate of return, and σ_i is the standard deviation of i.

The pairs selected by the first layer criteria where then scrutinized in-sample. The SR was computed for each pair in the validation sub-sample considering a risk-free rate given by annual rate of the Germany Government Bonds, $R_f = 3.604\%$, which was converted to a daily value to perform the calculations applying Equation (5.5).

As a thumb rule pairs with a $SR < 1$ were discarded. The remaining pairs,were 185 pairs out of 411 cointegrated pairs. On the GHE-formed pairs as none had an in-sample $SR > 1$, one considered $SR > 0.5$, obtaining 228 out of 792 pairs with an GHE lower than 0.45.

In order to reduce the number of pairs and obtain a better performance, one endogenized the minimum value of the acceptable Sharpe ratio value. This was accomplished by an in-sample grid search on the interval $[minSR, maxSR]$, where $maxSR$ and $minSR$ is the maximum and minimum Sharpe ratio obtained in-sample, respectively; 1 in the cointegration case and 0.5 in the GHE case. The grid included values of Sharpe ratio between minSR and maxSR, with a step of 0.01 between them. The minimum value of the Sharpe ratio obtained was equal to the maximum Sharpe ratio among the pairs, which led to only one pair being selected. A low number of pairs, even though profitable in-sample, may fail out-of-sample and compromise the strategy. A second search was performed, this time endogenizing the number of pairs with higher Sharpe ratio to be selected. The grid of the search contained values between 5 and the number of pairs with a Sharpe ratio higher than minSR, with a step of 5 between them. This time, it was determined that the optimal number of pairs with the best Sharpe ratio to be selected should be 20 highest SR out of the previous 185 for cointegration and 10 highest SR out of the previous 228 for the GHE selection. Figure 5.4 shows the variation of the in-sample return as a function of the number of pairs, suggesting that the average return peaked up for a lower numbers of pairs. In other words, the highest Sharpe ratios is followed by a decrease in the number of pairs. The decline was much faster in the GHE-formed pairs, while the cointegrated pairs declined less and at a certain number of pairs the return became almost constant.

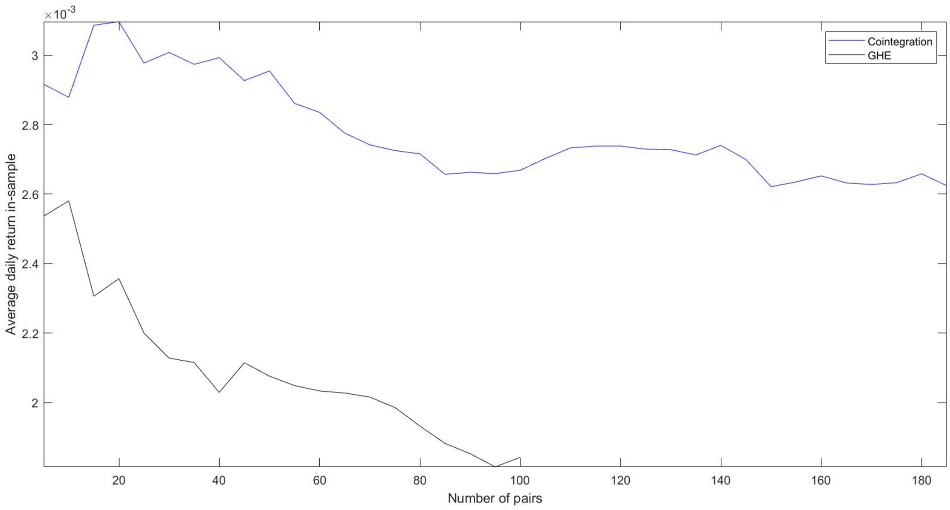

Figure 5.4 Average daily returns obtained in-sample for a varying number of higher in-sample Sharpe ratio for the cointegrated and GHE-formed pairs.

Criteria based on Ornstein-Uhlenbeck process parameters

Another second-layer selection criterion that is first proposed in this dissertation and not yet found in the literature is based on the parameters α and σ of the OU process. Its conceptualization comes from the observation of the results presented in Appendix A of Amer and Islam (2023) [1]. Often a higher value of mean reversion velocity, α, and also a higher volatility, σ, provide higher returns. Most probably, the rationale for this pattern is that a higher α, implies that the mean-reversion dynamics are faster, leading to a quicker profit realization, while a higher volatility, σ, potentially leads to higher spreads, which may also leads to higher profits.

The estimation of the OU parameters was done using the analytical expressions derived in Section 4.2 on the spread defined in Equation (3.1), where β is the cointegration coefficient or, in the Hurst exponent case, calculated using Equation (5.4). To select the best number of pairs with the highest parameter values, one tried to endogenize the minimum values of α and σ. For that, a grid search was performed, but the minimum values obtained were too high, which led to a low number of pairs.

As a compromise on return for a less risky strategy, one decided to endogenize the number of pairs with higher α and σ. A search grid was performed for both parameters, starting by selecting the 5 pairs with the highest value and increasing the number of pairs until all of the pairs were selected. Only the pairs with the highest values of α and σ are selected, having higher returns with some peaks but almost constant, followed by a decrease for a higher number of pairs. It is important to note that for the GHE-formed pairs, the returns as a function of the σ have a much faster decrease than for the cointegration-formed pairs. There is also a big difference between the returns variation with α and σ. This can be observed in Figure 5.5 and Figure 5.6, which led to select the number of pairs presented in Table 5.2.

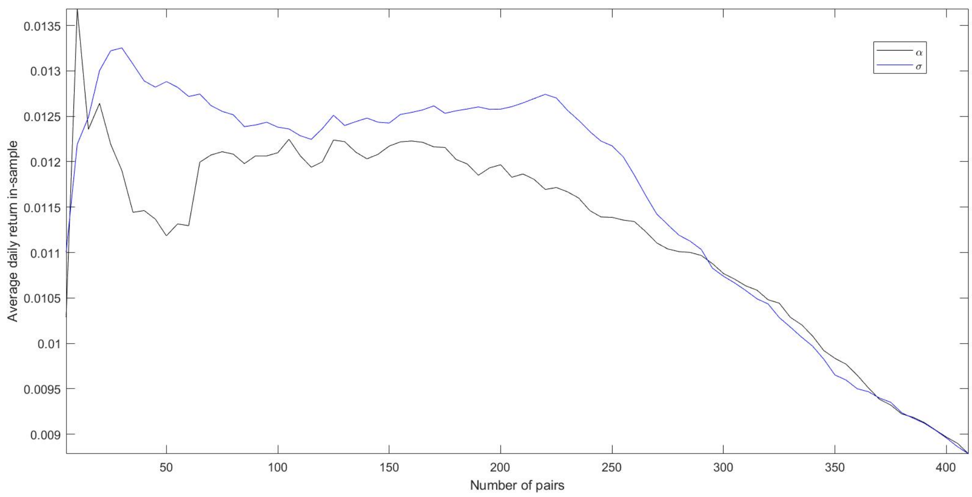

Figure 5.5 Average daily returns obtained in-sample for a varying α and σ in the cointegrated pairs.

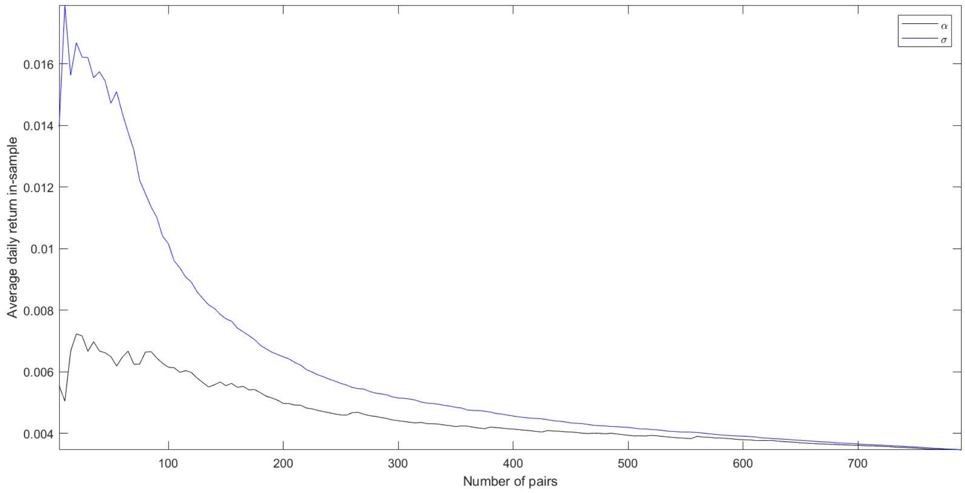

Figure 5.6 Average in-sample daily returns for a varying number of α and σ for the GHE-formed pairs.

Table 5.2 Limit values of the endogenized Ornstein-Uhlenbeck parameters

Pair formation	Parameter	Number of pairs	Minimum value selected	Minimum	Maximum
Cointegration	α	20	2.28	0.03	8.09
	σ	30	1.33	0.17	1.64
GHE	α	25	1.16	0.005	6.30
	σ	20	0.86	0.23	1.66

Notes: This table presents the best minimum value of the second layer criteria based on the Ornstein-Uhlenbeck process parameters α and σ, and the number of pairs that the selection leads to. The optimal values were obtained by a search grid in $[5, max]$ with a step of 5, where max is the total number of pairs (411 in the cointegration case, 792 in the GHE case). The maximum and minimum of these parameters for all pairs formed by cointegration or GHE, are also presented.

5.3 Threshold determination

The second stage has to do with thresholds determination, which define the values for opening and closing the pair position. Arguably, optimizing these parameters has a positive effect on the profitability of the trading strategies. The two methods used here to determine the thresholds are the Z-score (Caldeira and Moura, 2013 [5]) and the OU entropic method of Yoshikawa (2017) [32].

5.3.1 Z-score

The Z-score is a statistical measure between a data point and the mean of the data set. The formula for calculating the Z-score is $\frac{S_t - \mu_S}{\sigma_S}$, where S_t is the spread on day t and μ_S is the mean of the spread series, and σ_S is its standard deviation. The Z-score is employed to quantify the high and low historic values. Positive Z-score may indicate that an asset is overpriced relative to another asset, while a low negative Z-score may suggest underpricing.

One established entry and exit points for pair trading strategies based on the Z-score. For instance, a long (short) position is initiated in a moment of relative underpricing (overpricing) when the Z-score falls below (rises above) a certain negative (positive) threshold, anticipating a mean reversion where the spread increases (decreases). A more schematic and intuitive representations of this can be found in Figure 5.7.

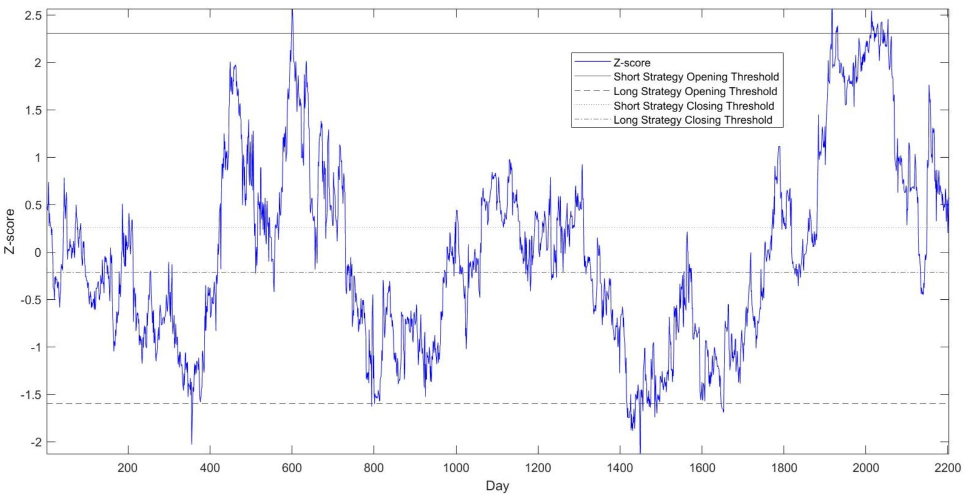

Figure 5.7 Z-score of spread between the assets 94, Artois, and 105, Beiersdorf, along with the long and short strategies opening and closing thresholds.

The determination of the thresholds was accomplished by a program that included four chained *for* cycles, choosing for each pair the combination of parameters that led to the highest return in-sample. The grid of searches was formed by percentages of the maximum spread value in the overvalued threshold case and a percentage of the minimum spread value in the undervalued threshold case. In the opening threshold case, the percentages values used were in a range from 10% to 90% in increments of 5%, whether the closing parameters ranged from 5% to 85% also in 5% increments. This led to 17 possibilities for each threshold and 83,521 different combinations of the four parameters for each pair. This way, one was able to endogenize the thresholds, and the returns obtained were very satisfactory. Table 5.3 shows the ranges of the thresholds obtained by this method for the 411 cointegration-formed pairs and for the 792 Hurst exponent-formed pairs.

5.3.2 Entropy Ornstein-Uhlenbeck method by Yoshikawa

The strategies constructed in this dissertation by this method were slightly different from the ones proposed by Yoshikawa (2017) [32]. Here, the long strategy starts when the spread is lower than the mean spread, parameter μ of the OU process, and is closed when the spread becomes higher than the optimal threshold b^*. The short strategy has its start when the spread is superior to b^* and closes when the spread reverts to the mean μ. The long strategy proposed has as an issue, it does not consider that

Table 5.3 Limit values of the endogenized thresholds of Z-score

Pair formation	Threshold	Minimum	Maximum
Cointegration	Short position opening	0.24	3.62
	Short position closing	0.13	1.88
	Long position opening	-4.03	-0.21
	Long position closing	-2.96	-0.12
GHE	Short position opening	0.22	2.45
	Short position closing	0.12	1.98
	Long position opening	-2.63	-0.20
	Long position closing	-1.87	-0.12

Notes: This table presents the endogenized thresholds of Z-score determined by a search grid between the percentages of the Z-score in a range from 10% to 90% in increments of 5% for the opening thresholds, whether the closing threshold ranged from 5% to 85% also in 5% increments. The minimum and maximum values are presented for each type of pair formation and position. GHE is an acronym for Generalized Hurst Exponent.

a deviation from the mean reverts, which is the mean-reversion principle, instead, it considers that the spread, when in the neighborhood of its mean, diverges from it, which is not based on a known principle. Besides, this strategy does not also take advantage of the relative underpricing of one of the assets. The long strategy is changed in this dissertation, while the short follows the Yoshikawa (2017) [32] proposal.

In order to account for the relative underpricing, one used a zero-centered spread, defined by $S'_t = S_t - \mu$, where S_t is defined in Equation (3.1), β is the cointegration coefficient, or, in the case of Hurst exponent selected pairs, is as defined in Equation (5.4), and μ is the mean spread. The threshold chosen to start the long strategy is the symmetric value of b^*, i.e., $-b^*$, and the closing point is the mean reverting point that, since the spread is centered, is zero. In Figure 5.8 it is possible to see a representation of a spread centered at zero, along with its b^* and $-b^*$, and the zero point, in this case the reverting point, in black. The dark blue arrow represents a short strategy, and the light blue a long strategy.

The value of b^* is obtained by solving Equation (4.2) whose zeros can be found using a non-parametric method since it is a complex equation. One chooses to apply the bisection method, a robust choice when derivatives are computationally resource-intensive. Unlike Newton's method, which relies on derivatives, the bisection method is derivative-free, making it suitable for this function where derivative information is hard to obtain. The integrals in the function were defined in Matlab by the pre-defined function *integral*. The calculation of b^* requires the OU process parameters values, which were estimated by the analytic expressions derived in Section 4.2. The expression also requires the discount rate, ρ, which was considered to be the risk-free rate as the annual rate of the German Government bonds, $\rho = 3.604\%$. The b^* values fluctuated between 0.14 and 2.90 for the cointegrated pairs and between 0.05 and 1.58 for the Hurst exponent pairs.

Having obtained the optimum threshold, one now obtains the thresholds b_t, which have the entropy as a penalty function in order to account for the model's uncertainty. The values vary along with the agent's confidence level in the probability measure, which is the confidence level λ. One considered three levels of confidence: $\lambda = \{0.001, 0.01, 0.1\}$. The values b_t are the solutions of Equation (4.6),

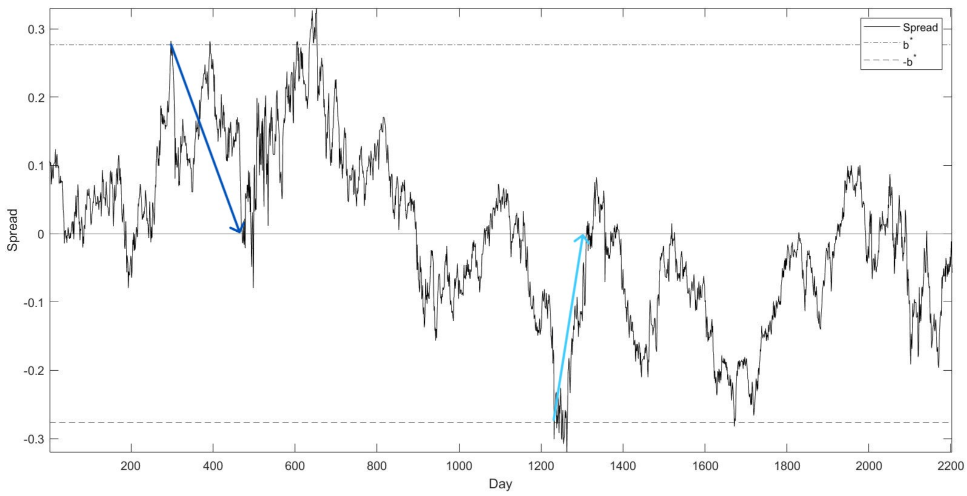

Figure 5.8 Zero-centered spread between the assets 58 and 79, Totalenergies and Dassault Aviation, respectively, along with the opening thresholds b^* and $-b^*$.

which was solved numerically by the bisection method for each time t, obtaining three series of values b_t for each pair, each with a different confidence level. Figure 5.9 shows a plot of the spread along with its respective b^* and b_t.

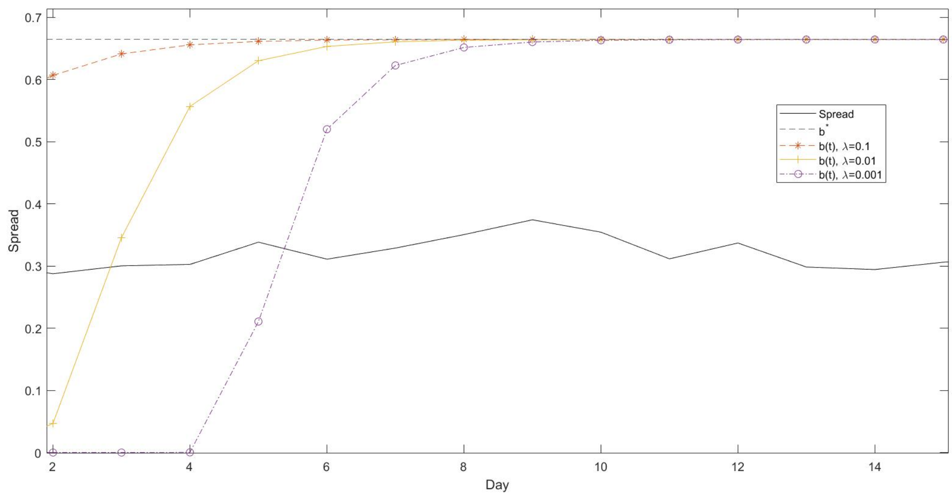

Figure 5.9 Zero-centered spread between the assets 91 and 97, Teleperformance and Siemens, respectively, along with the threshold b^* and the thresholds b_t for $\lambda = \{0.001, 0.01, 0.1\}$.

It is possible to see that b_t rapidly converges to the optimal threshold b^*, even though these thresholds create opportunities to open more strategies, as it is possible to see in the representation above. The strategies for these thresholds are constructed in the same way as the one with threshold b^*.

5.3.3 Performance measures

This section explains the metrics used to assess the performance of the strategies.

The average annual return of the portfolio, R_A, is obtained by:

$$R_A = 252 \frac{1}{N_T} \sum_{t=1}^{N_T} R_t, \tag{5.6}$$

where N_T is the number of days in the sample and R_t is the daily simple return on day t. The number of trading days per year is 252.

The Annualized compounded return (ACR) are also calculated in order to assess the cumulative effect that the gains or losses have on the capital over time. ACR is given by:

$$ACR = \left(\prod_{t=1}^{N_T} (1 + R_t) \right)^{\frac{252}{N_T}} - 1. \tag{5.7}$$

In order to better assess the performance of the pair trading strategies, one used some risk metrics. One of the metrics that can give a perception of how volatile these returns are is the annualized standard deviation (σ_A), given by:

$$\sigma_A = \sqrt{252} \sqrt{\frac{1}{N_T} \sum_{t=1}^{N_T} (R_t - \overline{R})^2}, \tag{5.8}$$

where $\overline{R}$ is the average daily return.

Another measure already introduced in Section 5.2.2 is the Sharpe ratio, which gives the risk-adjusted return and is given by Equation (5.5).

The Sortino ratio (SoR), a variation of the Sharpe ratio, utilises the asset's downside deviation, which is the standard deviation of negative portfolio returns, rather than the overall standard deviation of portfolio returns to distinguish between damaging and total volatility as a risk metric. This is significant because upside volatility is advantageous to investors. It is given by:

$$SoR = \frac{\overline{R}_p - R_f}{\sigma_D} \tag{5.9}$$

where $\overline{R}_p$ is the average daily return of the portfolio or strategy p, R_f is the risk free rate, and σ_D is the downside deviation of p.

The Maximum Drawdown (MDD) defines the maximum fall of an investment, given by the difference between the value of the lowest trough (R_t) and that of the highest peak before the trough (R_s). This is given by:

$$MDD = \frac{\underset{t \in [0, N_T]}{sup} \left[\underset{s \in [0,t]}{sup} R_s - R_t \right]}{R_s}. \tag{5.10}$$

The Value-at-Risk, VaR, measures the potential extent of financial losses of an investment over a given period of time. VaR is a tool used to quantify the risk exposure. One measures VaR by assessing the amount of potential loss, given the probability of occurrence, and the time frame. The probability used was 95%, which means that only 5% of the times the loss might surpass the VaR.

5.4 Main results

This section presents the results of the strategies. In Table 5.5, it is possible to find the results from strategies in which pairs were formed by cointegration. The strategy with the highest return was the one with the OU volatility parameter σ, as an additional criteria, selecting only the 30 most volatile pairs. This strategy yielded an annualized cumulative return (ACR) of 83.68%, more than 20% higher than what was obtained for the strategy with no additional criteria and the same threshold determination method. The strategy with no additional criteria obtained an ACR of 62.49%, higher than what was stated by Huck and Afawubo (2015) [15]. Returns can rise up to 60% for pairs formed by cointegration. For the other OU parameters additional criteria, the high α and high α combined with high σ, the lowest return is 45.70%, but are still lower than the strategy with no additional criteria. The use of the Sharpe ratio as an additional criteria proposed by Caldeira and Moura (2013) [5] was more profitable by 30%, than what was stated by the authors. Even though profitable, the Sharpe ratio as an additional criteria was not more profitable than the strategy with no additional criteria, C-Z. When determining the threshold by the Yoshikawa method, the return is much lower than when determining the thresholds by the endogenized Z-score. Again, selecting the pairs with high parameter σ of the OU process showed to be profitable, yielding profits from 23.30% up to 25.94%, while the same strategy with no additional criteria only presented a return from 7.91% to 8.60%. These values ware in accordance with Yoshikawa (2017) [32], who reports returns between 2.8% and 32.1%, and Amer and Islam (2023) [1], who report returns from 0.2% up to 25.2%. In this case, the other criteria based on the OU parameter did not have higher returns than the strategy with no additional criteria, C-Y. The Yoshikawa (2017) [32] proposal of using entropy as a penalty function proved profitable, existing at least a value of level of confidence, λ, that enhanced the profit. The C-OU-Y with high α and with high σ increased as the agents confidence level in the model probability measure decreased. The endogenized Z-score showed to be more profitable than the Yoshikawa (2017) [32] method for threshold determination. Besides being the most profitable C-OU-Z with high σ also showed to be well risk-adjusted, having the second highest Sharpe ratio and the third best Sortino ratio. Also, the VaR is only 4.45%. The best risk-adjusted strategy is the C-Z with a Sharpe ratio of 195.13%, and a Sortino ratio of 31.43%, the second highest. Also, the VaR of this strategy is 0%. In terms of risk, the strategies with thresholds defined by the Yoshikawa (2017) [32] method have low Sharpe and Sortino ratios and higher VaRs, being the only strategy with a Sharpe ratio close to 100%, the strategy with a high σ as an additional criteria. A possible conjecture about these lower returns is that the thresholds are lower than the ones of the endogenized Z-score, hence having a lower profit margin. The endogenized Z-score threshold for the opening of the short position fluctuate between 0.24 and 3.62 while the b^* for the same type of strategy goes from 0.14 to 2.90.

Table 5.6 shows the results for the pairs formed by GHE. The most profitable strategy was the one with high parameter σ of the OU process, with an ACR of 205.34%. The second most profitable strategy was the Sharpe ratio, which unlike cointegration, was more profitable than the strategy with no additional criteria, yielding 94.82%. The strategy, with no additional criteria and threshold determined by the endogenized Z-score, H-Z, had a ARC of 74.11%. The other strategies with OU parameters as additional criterion had good profits but still not as high as the high σ. In the case where the thresholds were defined by the Yoshikawa (2017) [32] method, the returns were much smaller, with the highest one being strategy H-OU-Y. Selecting the pairs with high σ, yielded a return of 64.24%. The second highest profit was obtained by the strategy with no additional criteria, H-Y, which profited a maximum of 7.00%. The strategy H-OU-Y with a combination of high α and high σ had negative returns, as low as -9.68%, which might be explained by its maximum drawdown of 111.10%, the highest one among all of the strategies. This strategy did not have a higher percentage of days with negative return, for all the strategies. This value was around 50%, but the positive daily returns were not sufficient to offset the negative daily returns. Once again, the threshold definition by endogenized Z-score showed to be more profitable than the Yoshikawa (2017) [32] method. The best risk-adjusted strategy was H-Z, with the highest Sharpe and Sortino ratios among all strategies and a VaR of 0%. The second highest Sharpe ratio was obtained in strategy H-OU-Z with the parameter α of the OU process as an additional criterion, with a VaR of 0%. The most profitable strategy, even though very volatile, with an annualized volatility of 126.88%, is still very well risk-adjusted, with a Sharpe ratio of 122.27% and a Sortino ratio of 34.28%. Strategies with thresholds determined by the Yoshikawa (2017) [32] were worse risk-adjusted than the ones with endogenized Z-score thresholds, having Sharpe ratios lower than one unity. In the GHE formed pairs, where the Yoshikawa (2017) [32] method was not yet implemented in the literature, the entropy as a penalty function did not show to be more profitable than the case where uncertainty is not considered ($\lambda \to \infty$). These results are better than the ones presented by Ramos-Requena (2017) [22] who obtained a return of 49.40% for GHE-formed pairs.

Overall, one concludes that the GHE is better than the cointegration method. This last method is only better in the Yoshikawa (2017) [32] threshold determination method. GHE only outperforms when considering the OU parameter σ as additional criteria. The superiority of GHE was already stated by Bui and Slepaczuk (2021) [4].

These profits, which are higher compared to most of the ones in the literature, can be explained by the illiquidity of some assets. The Portuguese market in 1990 was composed by several illiquid assets that did survive until the start of the test sample.

Looking at the time series of daily returns by pair, it was possible to see that a lot of these pairs in which one of the assets was illiquid had days where the returns were very high compared to pairs where both assets were liquid. Similar results were found by Broussard and Vaihekoski (2012) [3] that tried a pair trading strategy in Finland's 1990 market, which was also composed of some illiquid assets. The return obtained by the authors was 69.37% using a cointegration approach. The results obtained in this dissertation are still higher than this value, which can only be explained by the original points implemented that differ from the literature, such as the endogenization of parameters and the use of the OU parameters as additional criteria.

Another aspect tested was the exclusion of the crisis period that started with the Covid pandemic, considering only dates prior to March 1, 2020. The returns for the exact same strategies were reduced by at least 2%, which makes one conclude that the pair trading strategy is in fact market neutral, not being affected by the market downturn. In fact, the strategies did better in a period of market crisis, being able to capture the market's inefficiencies.

Comparing these results for the pair trading in Table 5.5 and in Table 5.6 to the ones presented in Table 5.4, one concludes that the pair trading strategy is not only more profitable than both benchmarks but also better risk-adjusted, having higher Sharpe and Sortino ratios in most strategies. Another important result is that when considering the data set without the illiquid assets, the *B&H* strategy return rises, unlike the pair trading strategies that actually profit from the market illiquidity.

Table 5.4 Results for the benchmark strategies

	$1/N$ portfolio	*B&H* portfolio
Average daily return	0.0382	0.0414
Average annualized return	9.62	10.44
Annualized volatility	16.05	18.05
Annualized compounded return	8.69	9.20
Sharpe Ratio	37.51	37.87
Maximum Drawdown	43.60	44.67
Sortino Ratio	3.64	3.67
Value-at-Risk	16.78	19.25

Notes: This table presents the results and risk metrics for the benchmark strategies, the equal-weighted portfolio $1/N$ and the buy-and-hold *B&H* portfolio. The strategies were applied to the 125 assets in the dataset through the test period, from May 1, 2015 to October 9, 2023. For each strategy it is presented the average daily return, as well as the average annuallized and annualized compounded returns. Risk metrics are also presented, such as annualized volatility, Sharpe ratio ,maximum drawdown, Sortino ratio and Value-at-Risk. All the results are presented in percentage.

Table 5.5 Results for the strategies with pairs formed by cointegration

Strategy	OU parameter criteria	λ	Average daily return	Average annualized return	Annualized volatility	Annualized compounded return	Sharpe Ratio	Maximum Drawdown	Sortino Ratio	Value-at-Risk
C-Z			0.2040	51.40	24.49	62.49	195.13	38.66	31.43	0
C-SR-Z			0.1540	38.82	33.46	39.67	105.24	64.20	13.07	16.22
C-OU-Z	High α		0.1736	43.75	35.52	45.70	113.02	73.28	13.89	14.68
	High σ		0.2817	70.99	45.87	83.68	146.92	66.25	18.71	4.45
	High α, σ		0.2684	67.64	89.40	61.85	71.63	26.32	34.77	79.42
C-OU-Y	High α	∞	0.0183	4.62	22.25	2.23	4.55	54.10	0.50	31.98
		0.1	0.0183	4.62	22.25	2.23	4.54	54.10	0.50	31.99
		0.01	0.0197	4.95	22.18	2.59	6.09	54.10	0.67	31.52
		0.001	0.0239	6.01	21.31	3.87	11.30	50.46	1.24	29.04
	High σ	∞	0.0955	24.08	25.16	23.30	81.37	60.55	9.33	17.31
		0.1	0.0954	24.03	25.08	23.27	81.46	60.55	9.34	17.22
		0.01	0.1030	25.96	23.86	26.05	93.72	51.92	11.16	13.28
		0.001	0.1017	25.64	22.85	25.94	96.44	47.68	11.44	11.94
	High α,σ	∞	0.0665	16.76	47.10	5.00	27.94	75.82	3.32	60.70
		0.1	0.0672	16.93	47.10	5.17	28.30	75.82	3.36	60.54
		0.01	0.0667	16.81	47.10	5.05	28.04	75.82	3.33	60.65
		0.001	0.0667	16.80	47.10	5.04	28.03	75.82	3.33	60.66
C-Y		∞	0.0367	9.26	14.47	8.60	39.08	30.08	5.19	14.54
		0.1	0.0358	9.01	14.19	8.38	38.11	29.74	5.06	14.54
		0.01	0.0337	8.49	13.50	7.91	36.20	28.62	4.82	13.71
		0.001	0.0337	8.49	13.44	7.91	36.33	28.20	4.85	13.62

Notes: This table contains the results of the pair trading strategies with pairs formed by the cointegration method. For each strategy it is presented the average daily return, as well as the average annualized and annualized compounded returns. Risk metrics are also presented, such as annualized volatility, Sharpe ratio, maximum drawdown, Sortino ratio and Value-at-Risk. The strategies were all implemented in the same time period, from May 1, 2015, to October , 2023. There were 125 assets from the Portuguese, Spanish, French and German markets. In the strategy line, the letters specify how the strategy was constructed. Letter C stands for pairs formed by cointegration, SR stands for the Sharpe ratio in-sample as an additional pair formation criteria, OU refers to the use of the Ornstein-Uhlenbeck parameters as an additional criteria for pair selection, Z refers to the strategy thresholds defined by Z-score, and Y refers to the thresholds defined by the Yoshikawa method. In the strategies with the Ornstein-Uhlenbeck parameters as additional , 3 different values of parameters were selected, high α, i.e. higher than the endogenized value, $\alpha = 2.28$, high σ, higher than the endogenized value, $\sigma = 1.33$. The case when both are high was also analysed. All the values presented are in percentage.

Table 5.6 Results for the strategies with pairs formed by GHE

Strategy	OU parameter criteria	λ	Average daily return	Average annualized return	Annualized volatility	Annualized compounded return	Sharpe Ratio	Maximum Drawdown	Sortino Ratio	Value-at-Risk
H-Z			0.2298	57.90	22.32	74.11	243.24	26.32	38.13	0
H-SR-Z			0.3839	96.74	94.78	94.82	98.26	91.03	20.82	59.17
H-OU-Z	High α		0.1961	49.42	28.56	57.44	160.44	39.28	19.83	0
	High σ		0.6299	158.74	126.88	205.34	122.27	73.54	34.28	49.96
	High α, σ		0.2559	64.49	46.94	73.21	129.71	55.75	19.59	12.72
H-OU-Y	High α	∞	0.0238	5.99	17.65	4.56	13.55	24.81	1.57	23.03
		0.1	0.0237	5.97	17.65	4.54	13.39	24.81	1.55	23.06
		0.01	0.0226	5.69	17.47	4.28	11.94	24.81	1.37	23.05
		0.001	0.0217	5.48	17.27	4.09	11.86	24.81	1.25	22.92
	High σ	∞	0.3167	79.80	100.25	64.24	76.01	67.00	22.87	85.09
		0.1	0.2986	75.24	92.59	62.08	77.37	67.00	22.35	77.05
		0.01	0.2897	72.99	89.97	59.99	77.13	67.00	21.48	74.99
		0.001	0.2897	72.99	89.97	59.99	77.13	67.00	21.48	74.99
	High α,σ	∞	-0.0284	-7.17	23.77	-9.49	-45.31	108.56	-4.91	46.26
		0.1	-0.0289	-7.28	23.75	-9.59	-45.85	109.59	-4.97	46.34
		0.01	-0.0293	-7.39	23.71	-9.68	-46.35	111.10	-5.03	46.39
		0.001	-0.0293	-7.39	23.71	-9.68	-46.35	111.10	-5.03	46.39
H-Y		∞	0.0337	8.50	19.44	6.91	25.20	29.76	3.45	23.48
		0.1	0.0340	8.57	19.39	7.00	25.63	29.66	3.50	23.31
		0.01	0.0334	8.41	19.21	6.86	25.01	29.58	3.41	23.19
		0.001	0.0333	8.34	19.20	6.83	24.90	29.62	3.40	23.20

Notes: This table contains the results of the pair trading strategies with pairs formed by generalized Hurst exponent. For each strategy is presented the average daily return, as well as the average annuallized and annualized compounded returns. Risk metrics are also presented, such as annualized volatility, Sharpe ratio, maximum drawdown, Sortino ratio and Value-at-Risk. The strategies were all implemented in the same time period, from May 1, 2015, to October 9, 2023. re are 125 assets, from which the pairs were formed, from the Portuguese, Spanish, French and German markets. In the strategy line, the letters specify how the strategy was constructed. Letter H stands for pairs formed by generalized Hurst exponent, SR stands for the Sharpe ratio in-sample as an additional pair formation criteria, OU refers to the use of the Ornstein-Uhlenbeck parameters as an additional criterion for pair selection, Z refers to the strategy thresholds defined by Z-score, and Y refers to the thresholds defined by the Yoshikawa method. In the strategies with the Ornstein-Uhlenbeck parameters as additional criterion, 3 different values of parameters were selected, high α, i.e. higher than the endogenized value, $\alpha = 1.16$, high σ, higher than the its endogenized value, $\sigma = 0.86$. The case where both are high was also analysed. All the values presented are in percentage.

Chapter 6

Conclusion

Pair trading is a type of statistical arbitrage strategy that has shown to be profitable in the literature. In order to create an efficient pair trading strategy, one must choose pairs that show a certain level of co-movement as well as a good choice of opening and closing thresholds. In the literature, several different strategies are formed by various pair formation methods and threshold definitions. This dissertation applies some of the best methods found in the literature, such as cointegration and generalized Hurst exponent (GHE), along with the Sharpe ratio as additional criteria. The threshold definition was performed by Z-score and the Yoshikawa method. This dissertation also corrects this method. Some original content is also presented in this dissertation, such as the endogenized values. Another original issue of this dissertation is the use of the Ornstein-Uhlenbeck (OU) parameters α and σ as additional criteria, testing how higher and lower values of mean-reversion velocity, α, and its combination with volatility, σ, affect the returns. The data set is formed by 125 stocks from the Portuguese, Spanish, French, and German markets. As the Portuguese market introduced illiquidity in the sample, this dissertation also brings more insight on how the pair trading strategy behaves under market illiquidity. The prime goal in this dissertation was to assess which combinations of methods to construct the pair trading strategy would perform better.

Through returns and other metrics, such as volatility, maximum drawdown, Sharpe ratio, Sortino ratio and Value-at-Risk, one has observed that the results are higher for the pairs formed by GHE. In the case where the Sharpe ratio is used as an additional criterion, this seems to be an enhancing method for the GHE-formed pair but not for the cointegrated ones, obtaining a lower profit than without the additional selection. Besides GHE-formed pairs being more profitable they are also better risk-adjusted, having higher Sharpe and Sortino ratios. The higher returns in both pair formations and both threshold definitions were obtained with the OU parameter σ as additional criteria. This parameters is the volatility of the spread, which means that a higher value of it might create more and bigger deviations from the mean, which in pair trading is an opportunity for profit. The best threshold definition was the Z-score with threshold endogenization, the Yoshikawa (2017) [32] method, especially without the OU parameter additional criteria. Besides having low profit, it also has low Sharpe ratios, lower than 100%. Another important aspect of the Yoshikawa (2017) [32] method is that it accounts for model uncertainty, considering the investors' level of confidence in the model, λ. How the confidence of the investor in the model affects the profit and risk of the pair trading strategy was also analyzed, concluding that in the analyzed pairs accounting for the uncertainty did not make the strategy more profitable nor better risk-adjusted in the GHE case while in the cointegrated pairs there is at least a value of level of confidence, λ, that enhanced the profit.

A comparison between the pair trading strategy and the benchmarks, equal-weighted portfolio and buy-and-hold strategy was also conducted, concluding that the pair trading beats the benchmarks in terms of return and risk. In order to assess the effects of the market downturn in the 2020 pandemic crisis it was performed a sample scaling down to only the period before the pandemic, which led to smaller profits. This supports the principle that pair trading strategy is market neutral performing better under market uncertainty.

Overall, one concludes that the best pair trading strategy is composed of GHE pair formation and Z-score endogenized parameters. Also, it is concluded that the OU parameter σ, as additional criteria, can enhance the profit no matter the method used. One also points to market illiquidity as one of the enablers for such high profits, showing that pair trading can still be profitable, if not more profitable under low liquidity conditions.

As future work one suggests a more in-depth study of the effects of the OU parameters as additional pair selection criteria in the returns of the pair trading strategy, since this is a new topic in literature that resulted in very satisfactory results.

Bibliography

[1] Amer, L. and Islam, T. U. (2023). An entropic approach for pair trading in PSX. *Entropy*, 25(3):494.

[2] Andrade, S., Di Pietro, V., and Seasholes, M. (2005). Understanding the profitability of pairs trading. *Unpublished working paper, UC Berkeley, Northwestern University*.

[3] Broussard, J. P. and Vaihekoski, M. (2012). Profitability of pairs trading strategy in an illiquid market with multiple share classes. *Journal of International Financial Markets, Institutions and Money*, 22(5):1188–1201.

[4] Bui, Q. and Ślepaczuk, R. (2022). Applying Hurst exponent in pair trading strategies on Nasdaq 100 index. *Physica A: Statistical Mechanics and its Applications*, 592:126784.

[5] Caldeira, J. and Moura, G. V. (2013). Selection of a portfolio of pairs based on cointegration: A statistical arbitrage strategy. *Available at SSRN 2196391*.

[6] Cropper, W. H. (1986). Rudolf Clausius and the road to entropy. *American Journal of Physics*, 54(12):1068–1074.

[7] DeMiguel, V., Garlappi, L., and Uppal, R. (2009). Optimal versus naive diversification: How inefficient is the 1/n portfolio strategy? *The Review of Financial Studies*, 22(5):1915–1953.

[8] Detlefsen, K. and Scandolo, G. (2005). onditional and dynamic convex risk measures. *Financ. Stoch.*, 9:539–561.

[9] Do, B., Faff, R., and Hamza, K. (2006). A new approach to modeling and estimation for pairs trading. *Financial Analysts Journal*, 1:87–99.

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

[11] Elliott, R. J., Van Der Hoek, J., and Malcolm, W. P. (2005). Pairs trading. *Quantitative Finance*, 5(3):271–276.

[12] Follmer, H. and Penner, I. (2006). Convex risk measures and the dynamics of their penalty functions. *Stat. Decis.*, 24:61–96.

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

[14] Hamilton, J. D. (1994). *Time series analysis*. Princeton University Press.

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

[16] Hurst, H. E. (1951). Long-term storage capacity of reservoirs. *Transactions of the American Society of Civil Engineers*, 116(1):770–799.

[17] Jacobs, H. and Weber, M. (2015). On the determinants of pairs trading profitability. *Journal of Financial Markets*, 23:75–97.

[18] Jaynes, E. T. (1965). Gibbs vs Boltzmann Entropies. *American Journal of Physics*, 33(5):391–398.

[19] Kullback, S. and Leibler, R. A. (1951). On information and sufficiency. *The Annals of Mathematical Statistics*, 22(1):79–86.

[20] Oksendal, B. (2013). *Stochastic differential equations: An introduction with applications*. Springer Science & Business Media.

[21] Rad, H., Low, R. K. Y., and Faff, R. (2016). The profitability of pairs trading strategies: distance, cointegration and copula methods. *Quantitative Finance*, 16(10):1541–1558.

[22] Ramos-Requena, J., Trinidad-Segovia, J., and Sánchez-Granero, M. (2017). Introducing Hurst exponent in pair trading. *Physica A: Statistical Mechanics and its Applications*, 488(2):39–45.

[23] Sarmento, S. M. and Horta, N. (2020). Enhancing a pairs trading strategy with the application of machine learning. *Expert Systems with Applications*, 158:113490.

[24] Shannon, C. E. and Weaver, W. (1949). *The mathematical theory of communication*. University of Illinois Press.

[25] Sharpe, W. F. (1998). The Sharpe ratio. *Streetwise–The Best of the Journal of Portfolio Management*, 3:169–85.

[26] Shiryaev, A. N. (2007). *Optimal stopping rules*, volume 8. Springer Science & Business Media.

[27] Simon, H. A. (1954). Spurious correlation: A causal interpretation. *Journal of the American Statistical Association*, 49(267):467–479.

[28] Suwarman, R. (2019). Real option valuation with stochastic interest rate and stochastic volatility. *Matematika*, 18(2).

[29] Uhlenbeck, G. E. and Ornstein, L. S. (1930). On the theory of the Brownian motion. *Physical Review*, 36(5):823–841.

[30] Vidyamurthy, G. (2004). *Pairs Trading: Quantitative methods and analysis*, volume 217. John Wiley & Sons.

[31] Wooldridge, J. M. (2012). *Introductory econometrics: A modern approach*. Southwestern College Publishing, 41st edition.

[32] Yoshikawa, D. (2017). An entropic approach for pair trading. *Entropy*, 19(7):320.

[33] Yu, F., Ching, W.-K., Wu, C., and Gu, J.-W. (2023). Optimal pairs trading strategies: A stochastic mean–variance approach. *Journal of Optimization Theory and Applications*, 196(1):36–55.

[34] Zivot, E. and Wang, J. (2006). *Modeling financial time series with S-PLUS*. Springer, 2nd edition.

Appendix A

Proof of the Yoshikawa method

In this appendix, one presents a detailed derivation of the method proposed by Yoshikawa (2017) [32], which uses an Ornstein-Uhlenbeck (OU) process to model the spread between the two assets. The proposed formulation is useful to calculate the optimal threshold of the strategy, b^*, which in a first state, is the output of the following optimal stopping problem:

$$v^0(x) = \sup_{\tau \in T} \quad \mathbb{E}_x[e^{-\rho\tau} Y_\tau], \tag{A.1}$$

where T is the set of all stopping times of the Wiener process, W, and ρ is the discount rate, and Y_t is an OU process as defined in Equation (3.11). It is also assumed that $\mathbb{E}_\tau < \infty$ for all $\tau \in T$.

In order to solve the optimal problem, one must consider that at time t, one has $x = Y_t$. Considering time t, the problem becomes the solution of $e^{-\rho t} v^0(x)$, which can be obtained applying Itô's formula, enunciated in Section 3.6. One has:

$$\begin{aligned} d(e^{-\rho t} v^0(x)) &= -\rho e^{-\rho t} v^0(x) dt + \frac{\partial e^{-\rho t} v^0(x)}{\partial x} dY_t + \frac{1}{2} \frac{\partial e^{-\rho t} v^0(x)}{\partial x^2} d\langle Y_t \rangle \\ &= \Big(-\rho e^{-\rho t} v^0(x) - e^{-\rho t} \alpha x v_x^0(x) + \frac{1}{2} e^{-\rho t} \sigma^2 v_{xx}^0(x) \Big) dt + e^{-\rho t} \sigma v_x^0(x) dW_t. \end{aligned} \tag{A.2}$$

To continue the derivation of the result, it is needed to apply the results of theorems 2.4 and 2.7 of Shiryaev (2007) [26]. These two theorems imply that the solution of this optimal problem requires the existence of a boundary b such that $v^0(b) = b$ if $x \geq b$, otherwise $v^0(x) > x$, and it also implies the martingale property, i.e., $-\rho v^0(x) - \alpha x v_x^0(x) + \frac{1}{2}\sigma^2 v_{xx}^0(x) = 0$. Using theorem 9.5 of Shiryaev (2007) [26], this problem holds the smooth fit condition, i.e., $v_x^0(b) = 1$ for $x = b$. This leads one to solve the following free-boundary problem:

$$-\alpha x v_x^0 + \frac{1}{2}\sigma^2 v_{xx}^0 = \rho v^0 \quad for \quad x < b, \tag{A.3}$$

$$v^0(b) = b \quad for \quad x \geq b, \tag{A.4}$$

$$v_x^0(b) = 1 \quad for \quad x = b. \tag{A.5}$$

From these expressions, one can derive the following expression:

$$v^0(x) = \frac{\sigma}{\sqrt{2\alpha}} \frac{\int_0^\infty u^{\frac{\rho}{\alpha}-1} e^{\frac{\sqrt{2\alpha}}{\sigma}ux - \frac{u^2}{2}} du}{\int_0^\infty u^{\frac{\rho}{\alpha}} e^{\frac{\sqrt{2\alpha}}{\sigma}ub - \frac{u^2}{2}} du}. \tag{A.6}$$

This result is useful to easily derive the expression for b^* that follows from $v^0(b^*) = b^*$, which is:

$$\frac{\sigma}{\sqrt{2\alpha}} \frac{\int_0^\infty u^{\frac{\rho}{\alpha}-1} e^{\frac{\sqrt{2\alpha}}{\sigma}ub^* - \frac{u^2}{2}} du}{\int_0^\infty u^{\frac{\rho}{\alpha}} e^{\frac{\sqrt{2\alpha}}{\sigma}ub^* - \frac{u^2}{2}} du} = b^*. \tag{A.7}$$

The derivation for the threshold b^* is concluded. The threshold b'_t, which uses entropy to account for model uncertainty, can be obtained as the solution to the following optimal problem:

$$\underset{\tau\in T}{sup}\ \underset{\mathbb{Q}\in\mathscr{Q}}{inf} \quad \mathbb{E}_x^{\mathbb{Q}}[e^{-\rho(\tau-t)}Y_\tau], \tag{A.8}$$

where $\mathscr{Q}$ is a class of probability measures on $(\Omega, \mathscr{F})$. This problem is called the maxmin expected utility, but as it often implies that the investor should not enter the market, a more flexible approach is considered, conditioning the problem with $x = Y_t$ at time t. Obtaining:

$$v(t,x) = \underset{\tau\in T}{sup} \quad \mathbb{E}_x^{\mathbb{Q}}[e^{-\rho(\tau-t)}Y_\tau], \tag{A.9}$$

where $\mathbb{Q}$ is the solution of

$$\underset{\mathbb{Q}\in\mathscr{Q}}{inf}\, \{\mathbb{E}_x^{\mathbb{Q}}[e^{-\rho(\tau-t)}Y_\tau] + \lambda e^{-\rho(\tau-t)} En_x[\mathbb{Q}|\mathbb{P}]\}, \tag{A.10}$$

where λ is a positive constant that reflects the confidence of the investor in the measure $\mathbb{P}$, when $\lambda \to \infty$ the investor has full confidence in the measure, while for $\lambda \to 0$ the investor has no confidence in the measure. $En[\mathbb{Q}|\mathbb{P}]$ is the relative entropy between the measures $\mathbb{Q}$ and $\mathbb{P}$, defined by Equation (3.14). The Expression (A.10) is given by:

$$\int \left(e^{-\rho(\tau-t)}Y_\tau + \lambda e^{-\rho(\tau-t)} ln\left(\frac{d\mathbb{Q}}{d\mathbb{P}}\right)\right) d\mathbb{Q} = \lambda e^{-\rho(\tau-t)} \int ln\left(e^{\frac{Y_\tau}{\lambda}} \frac{d\mathbb{Q}}{d\mathbb{P}}\right) d\mathbb{Q}. \tag{A.11}$$

The relative entropy is minimized when $\mathbb{Q} = \mathbb{P}$. Therefore, by applying thesame reasoning as in Follmer and Penner (2006) [12] and Detlefsen and Scandolo (2005) [8], optimally is attained by:

$$\left.\frac{d\mathbb{Q}}{d\mathbb{P}}\right|_{\mathscr{F}_t} = \frac{e^{-\frac{Y_t}{\lambda}}}{\mathbb{E}_x[e^{-\frac{Y_t}{\lambda}}]}. \tag{A.12}$$

As dY_t can be written in its integral form as $Y_t = e^{-\alpha t}(Y_0 + \int_0^t \sigma e^{\alpha s} dW_s)$, Expression (A.12) can be rewritten as:

$$\left.\frac{d\mathbb{Q}}{d\mathbb{P}}\right|_{\mathscr{F}_t} = \frac{e^{-\frac{1}{\lambda}\int_0^t \sigma e^{\alpha(s-t)} dW_s}}{\mathbb{E}_x[e^{-\frac{1}{\lambda}\int_0^t \sigma e^{\alpha(s-t)} dW_s}]}. \tag{A.13}$$

Applying the Girsanov theorem, enunciated in Section 3.7, the Wiener process under probability measure $\mathbb{Q}$, W^* is given by:

$$W_t^* = W_t + \frac{\sigma}{\lambda\alpha}(1 - e^{-\alpha t}), \tag{A.14}$$

that is,

$$dW_t^* = dW_t - \frac{\sigma}{\lambda\alpha} e^{-\alpha t}(-\alpha). \tag{A.15}$$

With this result the process Y_t becomes:

$$dY_t = -\left(\alpha Y_t + \frac{\sigma^2}{\lambda} e^{-\alpha t}\right) dt + \sigma dW_t^*. \tag{A.16}$$

Conditioning the optimal problem in Expression (A.9) on $x = Y_t$, one obtains:

$$v^1(t,x) = \underset{\tau \in T}{sup} \quad \mathbb{E}_x^{\mathbb{Q}}[e^{-\rho\tau} Y_\tau]. \tag{A.17}$$

In the case where $t \to \infty$, Y as defined in Equation (A.15) becomes as defined in Expression (A.1), from this it follows that $\underset{t\to\infty}{lim} \quad v^1(t,x) = v^0(x)$. Applying Itô's formula, Section 3.6, to solve $v(t,x) = e^{-\rho t} v^1(t,x)$, one obtains:

$$\begin{aligned} d(e^{-\rho t} v^1(t,x)) &= -\rho e^{-\rho t} v^1 dt + e^{-\rho t} v_t^1 dt + e^{-\rho t} v_x^1 \left(- \left(\alpha Y_t + \frac{\sigma^2}{\lambda} e^{-\alpha t}\right) dt + \sigma dW_t^* \right) + \frac{1}{2} e^{-\rho t} v_{xx}^1 d\langle Y \rangle_t \\ &= \left(-\rho e^{-\rho t} v^1 + e^{-\rho t} v_t^1 - e^{-\rho t} v_x^1 \left(\alpha Y_t + \frac{\sigma^2}{\lambda} e^{-\alpha t}\right) + \frac{1}{2} e^{-\rho t} \sigma^2 v_{xx}^1 \right) dt + e^{-\rho t} v_x^1 \sigma dW_t^*. \end{aligned} \tag{A.18}$$

Analogue to what was performed before, applying theorems 2.4 and 2.7 of Shiryaev (2007) [26] it is required a boundary b_t such that $v^1(t,b_t) = b_t$ if $x \geq b_t$, otherwise $v^1(t,x) > x$, and the martingale property holds, so that:

$$-\rho v^1 + v_t^1 - \left(\alpha Y_t + \frac{\sigma^2}{\lambda} e^{-\alpha t}\right) v_x^1 + \frac{1}{2}\sigma^2 v_{xx}^1 = 0. \tag{A.19}$$

Theorem 9.5 of Shiryaev (2007) [26] implies that $v_x^1(t,b_t) = b_t'$ for $x = b_t$. This way the free boundary problem becomes:

$$-\rho v^1 + v_t^1 - \left(\alpha Y_t + \frac{\sigma^2}{\lambda} e^{-\alpha t}\right) v_x^1 + \frac{1}{2}\sigma^2 v_{xx}^1 = 0 \quad for \quad x < b_t, \tag{A.20}$$

$$v^1(t,b_t) = b_t \quad for \quad x \geq b_t, \tag{A.21}$$

$$\frac{\partial v^1(t,b_t)}{\partial t} = b_t' \quad for \quad x = b_t. \tag{A.22}$$

Setting $v^1(t,x) = c_1 \int_0^\infty u^{\frac{\rho}{\alpha}-1} e^{\frac{\sqrt{2\alpha}}{\sigma}(x-g_t)u - \frac{u^2}{2}} du + c_2$, where $g_t = -\frac{\sigma^2}{\lambda} t e^{-\alpha t} + c_3 e^{-\alpha t}$, and applying these results to Equation (A.20) it follows that $c_2 = 0$ and noting that $\underset{t\to\infty}{lim} g_t = 0$, then one has

$\lim_{t\to\infty} v^1(t,x) = c_1 \int_0^\infty u^{\frac{\rho}{\alpha}-1} e^{\frac{\sqrt{2\alpha}}{\sigma}xu-\frac{u^2}{2}} du$. On the other hand, recall that:

$$\lim_{t\to\infty} v^1(t,x) = v^0(x) = \frac{\sigma}{\sqrt{2\alpha}} \frac{\int_0^\infty u^{\frac{\rho}{\alpha}-1} e^{\frac{\sqrt{2\alpha}}{\sigma}xu-\frac{u^2}{2}} du}{\int_0^\infty u^{\frac{\rho}{\alpha}} e^{\frac{\sqrt{2\alpha}}{\sigma}ub^*-\frac{u^2}{2}} du}. \tag{A.23}$$

This implies that $c_1 = \frac{\sigma}{\sqrt{2\alpha}} \frac{1}{\int_0^\infty u^{\frac{\rho}{\alpha}} e^{\frac{\sqrt{2\alpha}}{\sigma}ub^*-\frac{u^2}{2}} du}$. Substituting c_1 above:

$$v^1(t,x) = \frac{\sigma}{\sqrt{2\alpha}} \frac{\int_0^\infty u^{\frac{\rho}{\alpha}-1} e^{\frac{\sqrt{2\alpha}}{\sigma}(x-g_t)u-\frac{u^2}{2}} du}{\int_0^\infty u^{\frac{\rho}{\alpha}} e^{\frac{\sqrt{2\alpha}}{\sigma}ub^*-\frac{u^2}{2}} du}. \tag{A.24}$$

Equation (A.21) becomes:

$$v^1(t,b_t) = \frac{\sigma}{\sqrt{2\alpha}} \frac{\int_0^\infty u^{\frac{\rho}{\alpha}-1} e^{\frac{\sqrt{2\alpha}}{\sigma}(b_t-g_t)u-\frac{u^2}{2}} du}{\int_0^\infty u^{\frac{\rho}{\alpha}} e^{\frac{\sqrt{2\alpha}}{\sigma}ub^*-\frac{u^2}{2}} du} = b_t. \tag{A.25}$$

From Equation (A.22), which holds for $x = b_t$, one has:

$$\begin{aligned} v^1(t,b_t) + v^1_x(t,b_t)b'_t &= -g'(t) \frac{\int_0^\infty u^{\frac{\rho}{\alpha}} e^{\frac{\sqrt{2\alpha}}{\sigma}(b_t-g_t)u-\frac{u^2}{2}} du}{\int_0^\infty u^{\frac{\rho}{\alpha}} e^{\frac{\sqrt{2\alpha}}{\sigma}ub^*-\frac{u^2}{2}} du} + \frac{\int_0^\infty u^{\frac{\rho}{\alpha}} e^{\frac{\sqrt{2\alpha}}{\sigma}(b_t-g_t)u-\frac{u^2}{2}} du}{\int_0^\infty u^{\frac{\rho}{\alpha}} e^{\frac{\sqrt{2\alpha}}{\sigma}ub^*-\frac{u^2}{2}} du} b'_t \\ &= b'_t. \end{aligned} \tag{A.26}$$

Thus:

$$(b'_t - g'(t)) \frac{\int_0^\infty u^{\frac{\rho}{\alpha}} e^{\frac{\sqrt{2\alpha}}{\sigma}(b_t-g_t)u-\frac{u^2}{2}} du}{\int_0^\infty u^{\frac{\rho}{\alpha}} e^{\frac{\sqrt{2\alpha}}{\sigma}ub^*-\frac{u^2}{2}} du} = b''_t, \tag{A.27}$$

and

$$\begin{aligned} &(b'_t - g'(t)) \frac{-\frac{\rho/\alpha}{\rho/\alpha}\left(\frac{\sqrt{2\alpha}}{\sigma}(b_t-g_t)\right) \int_0^\infty u^{\frac{\rho}{\alpha}-1} e^{\frac{\sqrt{2\alpha}}{\sigma}(b_t-g_t)u-\frac{u^2}{2}} du}{\int_0^\infty u^{\frac{\rho}{\alpha}} e^{\frac{\sqrt{2\alpha}}{\sigma}ub^*-\frac{u^2}{2}} du} \\ &= b'_t. \end{aligned} \tag{A.28}$$

Substituting Equation (A.25) into Equation (A.28), one gets:

$$\frac{b'_t}{b'_t - g'(t)} = -\frac{2\alpha}{\sigma^2} \frac{\rho/\alpha}{1-\rho/\alpha}(b_t - g_t)b_t, \tag{A.29}$$

from which one obtains, with a correction to the derivation in Yoshikawa (2017) [32]:

$$ln(b_t) + \frac{\alpha}{\sigma^2}\frac{\rho}{\alpha-\rho}(g_t - b_t)^2 = ln(b^*) + \frac{\alpha}{\sigma^2}\frac{\rho}{\alpha-\rho}(b^*)^2, \quad g_t = -\frac{\sigma^2}{\lambda}te^{-\alpha t}, \tag{A.30}$$

recalling that when $t \to \infty$ it holds that $g_t \to 0$. This can be verified by remarking that, to have $x_t = x_\infty$, it must be $x_t' = 0$, that is, $[ln(b_t) + \frac{\alpha}{\sigma}\frac{\rho}{\alpha-\rho}(b_t - g_t)^2]' = 0$ Therefore, for $t \to 0$:

$$ln(b^*) + \frac{\alpha}{\sigma^2}\frac{\rho}{\alpha-\rho}(c_3 - b^*)^2 = ln(b^*) + \frac{\alpha}{\sigma^2}\frac{\rho}{\alpha-\rho}(b^*)^2, \tag{A.31}$$

since $g_t \underset{t\to 0}{\longrightarrow} c_3$. Equation (A.31) implies that $c_3 = 0$, finally obtaining:

$$ln(b_t) + \frac{\alpha}{\sigma^2}\frac{\rho}{\alpha-\rho}(g_t - b_t)^2 = ln(b^*) + \frac{\alpha}{\sigma^2}\frac{\rho}{\alpha-\rho}(b^*)^2. \tag{A.32}$$

Remark that, substituting $c_3 = 0$, $g_t = -\frac{\sigma^2}{\lambda}te^{-\alpha t}$. The derivation of the threshold b_t is concluded and with it the derivation of the Yoshikawa (2017) [32] method, recalling the correction to Theorem of Yoshikawa (2017) [32] in Equation (A.32).

Appendix B

List of stocks

Market	Number	Code	Company name	Sector
Portuguese	1	JMT	Jerónimo Martins	Personal Care, Drug & Grocery Stores
	2	BCP	Banco Comercial Portugês	Bank
	3	COR	Corticeira Amorim	Industrial Goods & Services
	4	GPA	Imobiliária Construtora Grão Pará	Travel & Leisure
	5	INA	Inapa	Industrial Goods & Services
	6	LIT	Litho Formas	Industrial Goods & Services
	7	CUR	Sociedade das Águas da Curia	Travel & Leisure
	8	SON	Sonae	Personal Care, Drug & Grocery Stores
	9	SNG	Sonagi	Real Estate
	10	SCT	Toyota Caetano	Automobiles & Parts
	11	VAF	Vista Alegre	Household Goods
	12	BPI	Banco BPI	Bank
	13	CIPN	Cipan	Pharmaceuticals
	14	COM	Compta	Technology
	15	CPA	Copam	Food Producers
	16	FEN	Fenalu	Financial Services
	17	SCO	SDC Investimentos	Construction & Materials
	18	ORE	Sociedade Comercial Orey Antunes	Financial Services
	19	SUCO	Sumol Compal	Food & Beverage
	20	TRAA	Transinsular	Industrial Transportation
Spanish	21	SAN	Banco Santander	Bank
	22	IBE	Iberdrola	Energy
	23	BBVA	BBV Argentaria	Bank
	24	TEF	Telefonica	Telecommunications
	25	ELE	Endesa	Energy
	26	NTGY	Naturgy Energy	Energy
	27	REP	Repsol	Energy
	28	ANA	Acciona	Infrastructures & Energy
	29	BKT	Bankinter	Bank
	30	IDR	Indra Sistemas	Techology

	31 MAP	Mapfre	Insurance
	32 NHH	NH Hotel	Travel & Leisure
	33 ACX	Acerinox	Industrial Goods & Services
	34 AMP	Amper	Energy & Techology
	35 AZK	Azkoyen	Techology
	36 EMP	Empresa Hidroelectrica Ribagorzana	Energy
	37 ALB	Corporation Financeira Alba	Financial Services
	38 MDF	Duro Felguera	Industrial Goods & Services
	39 ENC	Ence	Industrial Goods & Services
	40 ECR	Ercros	Industrial Goods & Services
	41 EZE	Grupo Ezentis	Infrastructures, Telecommunications & Energy
	42 COL	Inmobiliaria Colonial	Real Estate
Spanish	43 LGT	Lingotes Especiales	Automobiles & Parts
	44 NEA	Nicolas Correa	Industrial Goods & Services
	45 NYE	Nyesa Valores	Real Estate
	46 PSG	Prosegur	Security
	47 SCYR	Sacyr	Construction & Materials
	48 GPP	Service Point Solutions	Consulting & Techology
	49 TUB	Tubacex	Energy
	50 UBS	Urbas Guadahermosa	Real Estate
	51 VID	Vidrala	Industrial Goods & Services
	52 VIS	Viscofan	Industrial Goods & Services
	53 ADV	Adveo Group	Personal Care, Drug & Grocery Stores
	54 POP	Banco Popular Espanhol	Bank
	55 CGI	Compañía General de Inversiones	Financial Services
	56 LVMH	LVMH	Personal Goods
	57 OREP	L'Oréal	Personal Goods
	58 TTEF	Totalenergies	Energy
	59 SASY	Sanofi	Pharmaceuticals
	60 SCHN	Schneider Eletric	Electronic & Electrical Equipment
	61 ESLX	EssilorLuxottica	Health Care
	62 AIR	Air Liquide	Chemicals
	63 AXAF	Axa	Insurance
	64 SAF	Safran	Aerospace & Defense
	65 KER	Kering	Personal Goods
	66 SGEF	Vinci	Construction & Materials
	67 PERP	Pernod-Ricard	Beverages
French	68 DANO	Danone	Food Producers
	69 SGOB	Saint Gobain	Construction & Materials
	70 CAP	Capgemini	Technology
	71 TCFP	Thales	Aerospace & Defense
	72 SGE	Societe Generale	Banks
	73 MCL	Michelin	Automobiles & Parts
	74 PUB	Publicis Groupe	Media
	75 AIRF	Air France	Travel & Leisure
	76 BOLL	Bollore	Industrial Transportation
	77 BOUY	Bouygues	Construction & Materials
	78 CRFR	Carrefour	Personal Care, Drug & Grocery Stores
	79 AM	Dassault Aviation	Aerospace & Defense
	80 SDX	Sodexo	Restaurants & Bars

French	81	UBL	Unibail Rodamco Westfield	Real Estate
	82	AC	Accor	Travel & Leisure
	83	LOIM	Klepierre	Real Estate
	84	IMAF	Altarea	Real Estate
	85	ATO	Atos	Technology
	86	BIC	Bic	Nondurable Household Products
	87	CVO	Covivio	Real Estate
	88	FGR	Eiffage	Construction & Materials
	89	CFC	Gecina	Real Estate
	90	SEB	SEB	Household Goods & Home Construction
	91	TEP	Teleperformance	Industrial Support Services
	92	VIV	Vivendi	Media
	93	AREI	Altareit	Real Estate
	94	ART	Artois	Financial Services
	95	AURE	Aurea	Waste & Disposal Services
German	96	SAP	SAP	Technology
	97	SIE	Siemens	Industrial Goods & Services
	98	ALV	Allianz	Insurance
	99	BMW	BMW	Automobiles & Parts
	100	BAYN	BAYER	Pharmaceuticals
	101	MUV2	Meunchener Ruck	Insurance
	102	VOW	Volkswagen	Automobiles & Parts
	103	BAS	Basf	Chemicals
	104	DBK	Deutsche Bank	Bank
	105	BEI	Beiersdorf	Personal Care, Drug & Grocery Stores
	106	EOAN	E.ON Next	Energy
	107	RWE	RWE	Utilities
	108	LHA	Deutsche Lufthansa	Travel & Leisure
	109	CON	Continental	Automobiles & Parts
	110	EBK	EnBW Energie Baden-Württemberg	Energy
	111	CBK	Commerzbank	Bank
	112	HEI	Heidelberg Materials	Construction & Materials
	113	RHM	Rheinmetall	Aerospace & Defense
	114	BOSS	Hugo Boss	Personal Goods
	115	FPE3	Fuchs	Industrial Goods & Services
	116	HOT	Hochtief	Construction & Materials
	117	PAH	Porsche	Automobiles & Parts
	118	PUM	Puma	Personal Goods
	119	TUI	Tui	Travel &Leisure
	120	GBF	Bilfinger Berger	Industrial Engineering
	121	DEZ	Deutz	Industrial Engineering
	122	HNDG	Frosta	Food Producers
	123	G1A	GEA Group	Industrial Engineering
	124	SDF	K+S	Chemicals
	125	SZU	Suedzucker	Food Producers

Notes: This table presents a detailed description of the 125 assets that compose the data set. The table is divided by market, and presents the name of the asset, its code, its sector and a number that was attributed to every asset to be easily identified throughout the dissertation.