

Chapter 16

Statistical Arbitrage: An Approach from Econophysics



José Pedro Ramos-Requena , Antonio García Amate ,
and María de las Nieves López-García 

Abstract The statistical arbitrage strategy emerged in the mid-80s, developed by a group of scientists associated with Morgan Stanley's investment bank. This approach entails the choice of a pair of financial assets that have traditionally shown parallel movements. A profit can be achieved by establishing a long-short position on this pair when they move apart, resulting in gains upon their subsequent return to an average when closing the position. In this chapter, we are going to conduct a study on the application of this methodology by applying methods from the field of econophysics. To do this, we will examine the Hurst exponent method, which is one of the approaches that best fits to measure the mean reversion of financial series.

Keywords Statistical arbitrage · Pairs trading · Econophysics · Hurst exponent

16.1 Introduction

Statistical arbitrage is a financial strategy used to capitalize on market imbalances. This approach is rooted in the idea that past patterns will recur in the future, in contrast to fundamental investment strategies that aim to predict how economic factors influence stock prices.

In statistical arbitrage, the strategy involves taking long positions in assets with higher expected returns and short positions in assets with lower expected returns. This is done with the belief that these discrepancies are unrelated to overall market performance, making statistical arbitrage a market-neutral tactic. In essence, statistical arbitrage is a data-driven method designed to profit from market inefficiencies characterized by long-term deviations in stock prices observed in the past. Litterman (2004) outlines the philosophy of Goldman Sachs Asset Management, which assumes that even though markets may not be in equilibrium,

J. P. Ramos-Requena (✉) · A. García Amate · M. de las N. López-García
Department of Economics and Business, University of Almería, Almería, Spain
e-mail: jpramosre@ual.es; aga237@ual.es; mlg252@ual.es

they tend to gravitate towards balance over time, and traders seek to maximize gains from deviations from this equilibrium.

Concerning statistical arbitrage strategy, the most well-known application is undoubtedly Pairs Trading. This method identifies assets with similar historical price behaviors and aims to profit from short-term deviations from their long-term equilibrium (Lin and Tan, 2023). The pioneer in investigating this strategy was Nunzio Tartaglia's quantitative group at Morgan Stanley during the 1980s, and it remains an important statistical arbitrage technique used by hedge funds. Tartaglia's group found that certain assets exhibited correlations in their daily price movements, enabling them to create strategies to exploit market inefficiencies. Although the group initially achieved success in 1987, they faced challenges in 1989 and eventually disbanded. Nonetheless, the Pairs Trading strategy has gained popularity among individual traders, institutional investors, and hedge funds (Lo and MacKinlay, 1988; Khandani and Lo, 2011; Lo and MacKinlay, 1997; Gatev et al., 2006; Guidolin et al., 2009).

Pairs Trading is a strategy that identifies potential profit opportunities by statistically analyzing two assets simultaneously. What's particularly significant about Pairs Trading is its ability to remain market-neutral, which means it can generate profits regardless of whether the market is in a bearish or bullish phase. In a bear market, it involves taking short positions, while in a bull market, long positions are taken (Elliott and Bradrania, 2018). It's important to note that statistical arbitrage strategies are not considered true arbitrage strategies because they only yield a positive expected return within a finite time frame, rather than a guaranteed one. These strategies often involve executing numerous independent trades, each with a positive expected return, which helps mitigate overall strategy risk (Caldeira and Moura, 2013).

Market neutrality, as suggested by Khandani and Lo (2011), is attained by dividing stock returns into different quantiles and engaging in "winners versus losers" trading in a manner that is neutral concerning overall market exposure. In a more parametric approach, Poterba and Summers (1988) explore mean reversion through autoregressive models within the context of global stock markets. Zebedee and Kasch-Haroutounian (2009) focus on assessing the impact of pair trading at the microstructure level within the airline industry. Avellaneda and Lee (2010) incorporate principal component analysis or sector-based ETFs in their statistical arbitrage strategy. In all these scenarios, they model the residuals or unique components of a pairs portfolio as processes that revert to their mean.

One noteworthy application of this approach revolves around devising profitable trading strategies based on time series methods. This involves using a family of statistical arbitrage strategies that aim to exploit short-term instances of overvaluation and undervaluation among financial assets (Chen et al., 2019; Elliott et al., 2005). These situations are identified by examining the relative price adjustments between assets, rather than attempting to estimate an absolute value. If two assets exhibit similar characteristics, it is expected that their prices will follow a comparable pattern. Therefore, when an investor identifies two assets with akin characteristics, they can model the combination of their prices as a stable process

with a consistent average (Krauss et al., 2017). A significant deviation from this average indicates a mispricing between the two assets, where one is undervalued, the other is overvalued, or both scenarios are at play. In such instances, an investor would purchase the undervalued asset and take a short position on the overvalued one. Profit is realized when the mispricing between the assets corrects. The core assumption behind pair trading methods is the market's ability to rectify these mispricings.

The other critical aspect of Pairs Trading relates to deviations in the relative price of assets. To profit from these trades, these deviations should demonstrate a tendency to revert to a mean. However, mean reversion necessitates an underlying mechanism; pair trading wouldn't function if asset prices were genuinely random. The Law of One Price (LOP) asserts that assets with identical payoffs in all potential scenarios should have the same present value. Consequently, the price difference between closely substitutable assets should have a consistent, long-term equilibrium over time. Hendry and Juselius (2001) utilize this concept to illustrate that short-term deviations from these equivalent price conditions can create short-lived arbitrage opportunities based on the duration of the price deviation.

Implementing a Pairs Trading strategy presents two primary challenges. The first challenge involves identifying assets whose prices move in close correlation. The second challenge is developing a trading strategy that dictates when and in which asset to assume a long or short position. Traditional pair trading methods have typically relied on identifying trading pairs based on correlation and other non-parametric decision rules.

The primary emphasis of the leading studies focuses on adopting varied selection methods. A pivotal moment within the academic community was marked by the work of Gatev et al. (2006). This study introduces two distinct timeframes: the first for pairing assets and the second for actual trading. All asset pairs are categorized based on the total of squared differences (SSD) of normalized price series during the initial pairing phase. The subsequent months during the trading phase observe the best pairs identified with the lowest SSD value. A straightforward strategy based on distance is suggested: when the logarithmic price difference for a pair of assets significantly diverges from its average, it initiates a long position in the undervalued asset and a short position in the overvalued asset equally, forming a balanced portfolio of two assets. These positions are sustained until the price difference reverts to its average or until the end of the trading period.

Elliott et al. (2005) propose a Gaussian Markov chain model that reverts to the mean for the spread observed within Gaussian noise. Afterward, the model's predictions are compared to subsequent observations of the spread to make appropriate investment decisions. This approach has the potential to create wealth from any amount in financial markets that are perceived to be in a state of imbalance.

Vidyamurthy (2004) introduces the concept of cointegration for pair selection, a notion later implemented by Caldeira and Moura (2013). This study delves into the assessment of risk and returns of a portfolio composed of numerous pair transactions, all chosen based on cointegration.

Do et al. (2006) propose the modeling of the stochastic difference between two stocks, specifically the stochastic variance between the logarithms of their prices, within the context of a theoretical asset pricing model in equilibrium.

In addressing the challenge of multivariate Pairs trading, Huck (2009, 2010), employs a combined approach of machine learning and blended forecasting. This method is founded on future performance predictions using neural network models to attain anticipated variances.

Jacobs and Weber (2015) identify two potential drivers behind the profitability of pairs trading strategies: firstly, investors' potential overreactions to company-specific influences (termed the 'high attention hypothesis'); secondly, the gradual dissemination of common market impact information across select values (known as the 'limited attention hypothesis'). Their research findings predominantly support the concept of limited attention. This phenomenon becomes more pronounced, particularly in markets featuring a substantial array of tradable financial instruments, creating an 'information overload' leading to diminished visibility of individual pairs and further constraining the spread of common impact information. The limited attention hypothesis serves as empirical justification for techniques in multivariate pairs trading and statistical arbitrage, as advocated by Avellaneda and Lee (2010), Krauss and Stübinger (2017), Mrčela et al. (2017), Zhao et al. (2018, 2019), which focus on collections of values rather than individual pairs.

De Moura et al. (2016) sought to explain a pairs trading strategy by depicting cointegration and distance in a linear space model. Within this framework, the stochastic processes of cointegration and distance are represented based on mean reversion, acting as an imperceptible process within the linear space structure. Application of this approach in both the American and Brazilian stock markets resulted in superior returns compared to the overall market performance.

Hashemi et al. (2016) employ the ELECTRE-III technique to arrange all values based on the disparities among projected differences. From this classification method, the definitive portfolio is arranged. This entails opting for the highest-ranked values for a long position and the lower-ranked values for a short position. Both the long and short positions are initiated in identical measures, establishing a dollar-neutral investment portfolio.

Tadi et al. (2018) implemented a pair trading strategy with sets of stocks in the metallic minerals mining sector, analyzing 2015 price data. They contrasted the effectiveness of the pair trading strategy against a buy-and-hold approach, employing the pairs strategy as a subsequent examination. The findings suggested that, provided there is an optimally calibrated short-selling system, the pair trading strategy delivers superior returns compared to the buy-and-hold tactic.

Stübinger and Bredthauer (2017) analyzed the effectiveness of various pair trading strategies on the New York Stock Exchange spanning the years 1998 to 2015. The data they used had a high frequency and were scrutinized at a minute-by-minute level. Their findings highlighted that the most successful pair strategy yielded an annual return of 0.5050 with a Sharpe ratio of 8.14, encompassing transaction costs. Their study indicates that despite the declining profitability of pair trading strategies over time, they still present a viable choice for entering trading scenarios.

Clegg and Krauss (2018) explored the profitability associated with integrating the concept of partial cointegration within the pair trading system. They measured the parameters linked to partial cointegration through linear space representation and derived approximate coefficients using maximum likelihood estimation. Their results showcased that the strategy's profitability hovered around 12% per annum, accounting for transaction costs. This performance surpassed both the distance-based approach and market returns.

Elliott and Bradrania (2018) devised a dynamic model tailored for discrete pair trading. They considered several concealed states using the Markov switching approach within the distance process. Findings illustrate that the selection of the number of states influences the estimation of the model's parameters.

Zhao et al. (2018, 2019) introduced a technique to optimize the array of differences, assuming multivariate cointegration. Their objective was to identify those demonstrating stronger tendencies to revert to the mean and possessing higher variance to amplify profit potential. However, within the multivariate cointegration framework, the assumption is that price series (or logarithmic prices) maintain a stable linear relationship. This assumption delivers refined mathematical attributes but confines the admissible structures of multivariate movement.

Lu et al. (2022) devised a two-phase strategy for optimizing pair trading that incorporated a pair strategy to identify structural changes using machine learning techniques. The initial phase encompassed a hybrid model that utilized features from both frequency and time domains to detect structural changes, while the second phase implemented pair trading, identifying significant risks such as structural changes and market closures by employing a novel reinforcement learning model.

Kim et al. (2022) introduced HDRL-Trader, a hybrid deep reinforcement learning technique for pair trading. This method utilizes two separate reinforcement learning networks: one to determine trading actions and another to establish stop-loss limits. Additionally, HDRL-Trader integrates inventive methods like dimensionality reduction, clustering, regression, behavior replication, prioritized experience play-back, and dynamic delay within its structure. The performance of HDRL-Trader is compared against state-of-the-art reinforcement learning approaches for pair trading (P-DDQN, PTDQN, and P-Trader).

Han et al. (2023) devised a pair trading approach using unsupervised learning that identified pairs by amalgamating corporate features (foundational details) with pricing data. These company characteristics were leveraged to spot pairs and significantly boost the effectiveness of the pair trading strategy.

Keshavarz Haddad and Talebi (2023) contended that market professionals and speculators seek to capitalize on market disparities using arbitrage strategies. While some traders use financial and fundamental stock information, others profit from risk-hedging tactics and swing trading. Pair trading, a subset of statistical arbitrage, offers relatively secure risk-free gains. This study aimed to create a hypothetical portfolio of stock pairs exhibiting a notable price relationship on the Toronto Stock Exchange (TSX). Assessing strategies between January 2017 to June 2020, they compared the profitability of distance methods, cointegration, and Copula models.

The study revealed that Copula trading yielded the most substantial profits. Despite the complexities during the COVID-19 period, these strategies remained resilient.

Šarić et al. (2023) presented a new portfolio construction method based on preference relation graphs, which can resolve contradictory signals from pair trading across multiple sets of securities. The proposed approach allows for the joint exploitation of arbitrage opportunities among a large number of assets. Experimental results using three decades of historical returns from approximately 500 stocks in the S&P 500 index show that portfolios based on preference relations yield robust returns even with high transaction costs, and their performance improves with an increasing number of assets considered.

Brunetti and De Luca (2023) conducted an assessment of the effectiveness of the index options market by employing a statistical arbitrage approach, utilizing information sourced from five index options markets within the Euro Area.

This chapter, we are going to study the new field of study called econophysics applying it to the statistical arbitrage strategy. After this introduction, in Sect. 16.2, we are going to define the main selection methods that have been used so far in the statistical arbitrage strategy. Section 16.3 will be dedicated to the application of econophysics in the Pairs Trading strategy through the methods of the Hurst exponent and the HP approach. In Sect. 16.4, we define the methodology to implement the statistical arbitrage strategy, and finally, in Sect. 16.5, we will define the main conclusions of this chapter.

16.2 Pair Selection Methodology

This section explains in detail the main methods defined in pairs trading strategies to select pairs of securities to form a portfolio.

16.2.1 Distance Approach

This method was pioneered by Gatev et al. (2006) and has been one of the most widely used in the scientific literature (Andrade et al., 2005; Engelberg et al., 2009; Perlin, 2009; Do and Faff, 2010; Bowen et al., 2010; Do and Faff, 2012). This methodology defines co-movement of two assets as the squared distance of the normalised price series.

Through this approach, the smaller the distance between the prices of two stocks, the greater the co-movement.

Once the price of these shares has been normalised, considering two assets A and B, we could calculate their distance through the following expression:

$$d_{A,B}^2 = ||P_A - P_B||^2 \quad (16.1)$$

Those pairs with the shortest distance will be selected and subsequently monitored during the performance period.

Although this approach is very simple, it varies from study to study and its results are very difficult to compare.

From a statistical point of view, the distance method is equivalent to the maximum correlation criterion (Liew and Wu, 2013).

Among the criticisms that have been made of this method, it is worth mentioning the paradox that exists between the criterion of choosing pairs in the formation period and the objective of generating high returns in the performance period. The paradox lies in the fact that those pairs of stocks that are closest (those with the shortest distance) are chosen and are expected to be the ones most likely to deviate from the mean in the future and subsequently converge (Do and Faff, 2010). These authors propose to take into account the number of times the distance between the normalised stock prices crosses the mean to select the pairs (Do and Faff, 2012).

16.2.2 Correlation Method

Of the different methods that exist to measure correlation, in this case, we are going to define Pearson's correlation coefficient and Spearman's correlation coefficient, since they are the most widely used in the different studies applied in finance.

16.2.2.1 Pearsons Correlation Coefficient

This method consists of selecting those pairs that are highly correlated, i.e., that have a coefficient close to 1.

The different potential pairs should be analysed and the correlation coefficient between each of them calculated.

Preston (2005) argued that stocks or indices that are good candidates for selection for the Pairs Trading strategy should have some measurable relationship. Ideally, they should have a positive correlation.

In the following, we will explain how the Correlation approach works as applied to the Pairs Trading strategy.

First, we will calculate Pearson's correlation coefficient:

$$\rho_{A,B} = \frac{Cov(AB)}{\sigma_A * \sigma_B} \quad (16.2)$$

where:

- $\rho_{A,B}$ is the Pearson correlation coefficient for A and B shares.
- $Cov(AB)$, is the covariance between the returns of stocks A and B .
- σ_A is the standard deviation of the return on stock A .
- σ_B is the standard deviation of the return of stock B .

Once the pair of stocks with the highest correlation has been obtained, Wong (2010) proposes drawing up a new series of data obtained from a ratio of the prices of both stocks, which for reasons of simplification divides the higher value by the lower value to obtain results greater than unity. This ratio represents the average behaviour between the prices of both stocks and, therefore, upward or downward deviations from this value may represent investment opportunities.

16.2.2.2 Spearman Correlation Coefficient

This coefficient was proposed by Charles Spearman and is a non-parametric, distribution-free rank measure of the strength of the association between two variables. The Spearman correlation coefficient, often called Spearman's rank correlation, quantifies the association between two variables. In contrast to Pearson's correlation coefficient, which examines the linear connection between two variables, the Spearman correlation coefficient is employed to assess the ranking relationship or the extent of monotonicity in the link between the two variables.

This method attempts to describe the relationship of two variables in an arbitrary monotonic function without making any assumptions about the frequency distribution of the variables (Hauke and Kossowski, 2011).

For a sample of size n , the random variables A_i , B_i are converted into ranks rg_{A_i} , rg_{B_i} and $r(s)$ is considered, calculated from:

$$r_s = \rho_{rg_A, rg_B} = \frac{cov(rg_A, rg_B)}{\sigma_{rg_A} * \sigma_{rg_B}} \quad (16.3)$$

where:

- ρ_{rg_A, rg_B} is Pearson's correlation coefficient of the range variables.
- $cov(rg_A, rg_B)$ describes the covariance of range variables.
- σ_{rg_A} and σ_{rg_B} reflects the standard deviations of the range variables.

16.2.3 Cointegration Approach

The concept of cointegration was pioneered by Engle and Granger in their work (Engle and Granger, 1987). It's a statistical technique that captures various forms of co-movement dynamics. In this context, we are not referring to the co-movement of returns but rather the co-movement of stock prices (Alexander et al., 2002a). Some studies have suggested that correlation and cointegration methodologies complement each other. Cointegration focuses on the relationship between time series themselves, while correlation primarily deals with the relationship between the derivatives or changes in the series (Alexander et al., 2002b).

The cointegration method stands as one of the most commonly employed techniques in studies related to Pairs Trading. Burgess (1999) was among the first authors to apply this method, utilizing data from the FTSE 100 index. In the realm of Pairs Trading strategy research, one of the most frequently referenced works is the theoretical study conducted by Vidyamurthy (2004).

We can state that the Cointegration method is a statistical relationship in which two integrated series of order d can be linearly combined to create a third integrated series of order $d - b$, where $b > 0$. The most general and widespread case occurs when $d = b = 1$ (Do et al., 2006).

If we consider P_t^A and P_t^B the prices of stocks A and B , respectively, if both prices are individually non-stationary, there is a parameter γ such that the following equation makes it a stationary process:

$$P_t^A - \gamma * P_t^B = \mu + \epsilon_t \quad (16.4)$$

where:

- μ is the mean of the Cointegration model
- ϵ_t is the residual
- γ is the cointegration coefficient

The regression parameters will be estimated using the Ordinary Least Squares (OLS) method. To verify that the two stocks are cointegrated, we must verify whether the residual ϵ_t is stationary or not, for this we will use the Augmented Dickey Fuller test.

Thus, the strategy for trading with this method consists of calculating a level of deviation from the positive and negative mean and, once reached, buying or selling positions can be taken on the portfolio of A and B shares. A given level of deviation can be defined based on the historical behaviour of the series that maximises the result.

16.2.4 Other Methods

Other methods have been used to analyse the Paris Trading strategy, which have been used less frequently or have only recently been developed.

Elliott et al. (2005) develop a novel Gaussian Markov chain-based mean-reverting mean-reversion approach for the observed spread with Gaussian noise. The obtained values are compared with the posterior observations of the spread to decide on appropriate investment decisions. This approach can be applied to generate wealth in out-of-equilibrium financial markets.

The copula method applied to the statistical arbitrage strategy consists of searching for the optimal copula between the returns of two assets and identifying the different positions between pairs of securities. For this purpose, marginal

distributions, the copula function and conditional probability distribution functions should be used (Liew and Wu, 2013).

Compared to classical methods, this method provides richer information, such as the form and nature of the dependence between pairs of assets (Ferreira, 2008). This is because copulas measure dependencies of upper and lower tails of different scope, in an environment that considers linear and nonlinear relationships. Copulas have the property that they are treated invariant under strictly monotonic transformations of the random variables, that is, the same copula will be obtained regardless of whether we use price series or price logarithm series (Hu, 2006).

Regarding the way to calculate copulas to select pairs, different methods have been used to calculate them. Liew and Wu (2013) and Ferreira (2008) employ parametric distribution functions, on the other hand, Stander et al. (2013) discuss parametric and non-parametric approaches to obtain the marginal distributions. In the case of Xie et al. (2015), he develops a new method using the copula modeling technique. This approach allows estimating separately the marginal distributions of asset returns and their joint dependence structure. This provides a more robust estimate of the joint distribution of stock returns. They compare the results obtained through this approach against the distance method, in which they conclude that greater benefits are obtained with the distance method.

Stübinger et al. (2018) develop a multivariate pair trading strategy based on vine copulas, which is a very flexible method for modeling linear and nonlinear multivariate dependence. Da Silva et al. (2018) propose a mixed model based on copulas with the objective of obtaining linear and nonlinear associations and, at the same time, covering a wider range of possible dependence structures. Their main objective is to evaluate and develop a more complex strategy, taking advantage of any anomaly in the market to discover a relationship or pattern, with a higher profitability.

With technological progress, models applying machine learning have been introduced to identify non-linear structures in financial markets. In the field of the application of this methodology in the field of statistical strategy we can highlight Huck (2009, 2010) applying Elman neural network ensembles and ELECTRE III, a multi-criteria decision method, to statistical arbitrage strategy; Takeuchi and Lee (2013) develop an improved momentum strategy for US CRSP stocks, for this they use deep neural networks, to rank the probability of outperforming the cross-sectional mean return of all stocks during a trend month; Moritz and Zimmermann (2014) employ random forests for US CRSP stocks to implement the statistical arbitrage strategy based on deep conditional portfolio classes; Dixon et al. (2015) apply a similar strategy to Takeuchi and Lee (2013) applied to intraday data, with 5-minute pooled returns.

Krauss et al. (2017) develop and apply deep neural networks (DNN), gradient boosted trees (GBT), random forests (RAF) in the field of statistical arbitrage strategy and analyse the effectiveness of deep neural networks (DNN), gradient boosted trees (GBT), random forests (RAF) and various sets of these methods in the context of statistical arbitrage applied to the S&P 500 market. Fischer (2018) use short term memory networks (LSTM) which allows sequence learning to be

generated, this method is not typically used for time series forecasting, although they are valid for this domain. Fischer et al. (2019) apply machine learning for a statistical arbitrage strategy in the cryptocurrency market by training a random forest on the returns of 40 cryptocurrencies, with the objective of predicting whether a cryptocurrency will outperform the median cross-sectional of the 40 coins in the next 120 minutes. Chan (2023) studies the performance of the Pairs Trading strategy through neural network techniques applied to the oil and gas industry.

16.3 Application of Econophysics in Pairs Trading Strategy

Econophysics is a relatively young and continually developing field. One of its notable contributions to finance is the fractal market hypothesis, which challenges the conventional efficient market hypothesis. This fractal market hypothesis has paved the way for the emergence of fresh research and models (López-García and Ramos-Requena, 2019).

Econophysics is a relatively recent area of investigation that employs principles and techniques from statistical physics to comprehend and address economic issues, particularly those characterized by stochastic elements and nonlinear dynamics.

This emerging field of inquiry originated from the curiosity of physicists to provide mathematical explanations for phenomena in the social sciences. It is believed that this field started taking shape in the 1990s at the Santa Fe Institute in New Mexico, a center specializing in the examination of complex systems. While the term “econophysics” is of more recent origin, having been coined by the scientist Eugene Stanley at Boston University (Stanley et al., 1999).

Econophysics aims to bridge the gap between economics and observable realities by establishing connections with theoretical models. While economics encompasses a wide array of issues, econophysicists notably concentrate their efforts, particularly in finance, where they delve into the dynamics of market prices. The swift rise of econophysics in these studies can be attributed to the abundant financial data available from the 1980s onward, enabling the application of various mathematical models from physics, like examining memory within the financial market.

This section primarily focuses on exploring market long-term memory and consequently identifying mean reversion by employing the Hurst exponent method. The examination of memory processes has become prevalent across diverse fields of knowledge in recent years, spanning from hydrogeology to meteorology and extending to the study of internet traffic patterns and even heart rate patterns (Hurst, 1951; Mandelbrot and Wallis, 1969; Movahed and Hermanis, 2008; Peng et al., 1992).

In the realm of social sciences, especially in economics, the investigation of memory processes piqued researchers’ interest in the 1970s (Mandelbrot, 1971, 1972). Mandelbrot (1997) spearheaded the incorporation of the Hurst exponent into the analysis of financial time series.

Over the past few decades, the analysis of memory processes has expanded from macroeconomics to finance, where it has gained significant traction. The finance sector's interest in memory process analysis stems from its potential as a valid means to test the Efficient Market Hypothesis. However, it's important to note that drawing conclusions from this approach (Sánchez-Granero et al., 2012), involves considerations beyond its seemingly straightforward premise.

Next, we will develop the two methods that have utilized the Hurst exponent as a means of selecting pairs in the statistical arbitrage strategy.

16.3.1 Hurst Exponent

The concept of the Hurst exponent can be traced back to 1951 (Hurst, 1951) when the hydrologist Harold Edwin Hurst proposed a methodology for optimizing the storage capacity of a reservoir intended to regulate the natural inflows of the Nile River. This approach considered cyclical trends in flow, droughts, and, by extension, floods. Hurst's method is founded on the notion that many natural phenomena exhibit behavior that can be characterized as a biased random process, where there is a form of memory between observations. This approach can be described as a measure of the independence of a series and serves as a tool for identifying fractal behavior. Hurst's observations revealed that most of the natural phenomena he studied follow a biased random process, with long-term memory between observations. In other words, what occurs in one period will influence future events (Delgado and Delgado, 2011).

The Hurst exponent, sometimes referred to as the Hurst dependence index, long-term dependence index, or long-term memory index, is a metric that quantifies the relative inclination of a time series to revert to its mean or exhibit clustering in a particular direction (Kleinow, 2002). When the Hurst exponent (H) falls within the range of 0.5 to 1, it signifies that the time series exhibits positive long-term autocorrelation. In practical terms, this means that a high value in the series is more likely to be followed by another high value, and over a period of time, values tend to stay on the high side. On the other hand, if the Hurst exponent is within the range of 0 to 0.5, it indicates that the series experiences long-term alternation between high and low values. In this scenario, a single high value is likely to be succeeded by a low value, and the series tends to oscillate between high and low values over time. It's worth noting that a value of $H = 0.5$ characterizes a Brownian motion, which suggests a random and unpredictable series with no long-term autocorrelation or directional bias.

In the field of economics, the study of memory processes has expanded over recent decades. It initially emerged in macroeconomics, as seen in works by Diebold and Rudebusch (1989), Hassler (1994), Hassler and Wolters (1995), and Conniffe and Spencer (2000). However, it has gained significant traction in finance, where it has found the most widespread acceptance. The financial sector's interest in memory processes is largely driven by their potential to serve as a relatively straightforward

means to test the Efficient Market Hypothesis. In essence, the significance of memory processes in financial markets lies in their ability to challenge the Efficient Market Hypothesis. If a particular security exhibits long-term memory, it suggests cyclical behavior and deviates from a purely random process, directly undermining the efficiency hypothesis. Nonetheless, it's important to note, as indicated in research such as Sánchez et al. (2015), that drawing such conclusions is not as straightforward as some authors may suggest, as there are other aspects that require consideration.

In the field of statistical arbitrage, the pioneers in using this methodology to select pairs are Ramos-Requena et al. (2017). Since one of the main characteristics of statistical arbitrage is mean reversion, to fulfil this property they select those pairs with a $H < 0.5$, that is, those pairs with an antipersistent relationship. To select the pairs they use the Generalised Hurst Exponent (GHE), the Generalised Hurst Exponent (Barabási and Vicsek, 1991), since they indicate that it is the simplest method of calculation and allows them to obtain more reliable results.

The GHE is one of the most popular methods for calculating the Hurst exponent. It is a generalisation of the classical approach provided by Hurst (1951) and relates to the scaling behaviour of some statistical properties of a time series. It is considered a powerful tool to detect multifractality by scaling moments of order q of the distribution of increments. This scaling property is determined by an exponent H which is usually connected to the long-term statistical dependence of the time series.

In particular, it has been verified that these statistical properties of the time series scale with both the observation period (T) and the resolution of the time window. For this purpose, the following statistic is considered:

$$K_q(\tau) = \frac{\langle |X(t + \tau) - X(t)|^q \rangle}{\langle |X(t)|^q \rangle} \quad (16.5)$$

where τ can vary between 1 and τ_{max} . So that τ_{max} is generally chosen as a quarter of the length of the series and $\langle \cdot \rangle$ describes the sample average over the time window.

Therefore, the GHE is defined from the scaling behaviour of the statistic $K_q(\tau)$ given by the power law:

$$K_q(\tau) \propto \tau^{qH(q)}. \quad (16.6)$$

The GHE is determined by calculating an average of a set of values derived from the expression found in Eq. (16.6). These values are computed for various values of τ , as detailed in studies such as Di Matteo and Aste (2005), Barunik and Kristoufek (2010).

$$K_q(\tau) = \sum_{t=1}^{T-\tau} \frac{|X(t + \tau) - X(t)|^q}{T - \tau + 1} \quad (16.7)$$

For a time series $X(t)$ of length T , which also varies as predicted by (16.6). We should note that all the information about the scaling properties of the series is contained in the exponent $H(q)$ which makes the GHE-based analysis quite simple.

Since the development by Ramos-Requena et al. (2017), different studies have been applied to analyse the statistical arbitrage strategy by applying the Hurst Exponent approach. Bui and Ślepaczuk (2022) seek an alternative approach to pair selection by applying the Generalised Hurst exponent, taking 103 stocks of the Nasdaq 100 stock price index, comparing it with the cointegration and correlation methods. They conclude that it is an efficient market, and the results obtained are superior to the methods compared. Sánchez-Granero et al. (2020) employ this approach with the main objective of whether Latin American stock markets are efficient. To do so, they employ the statistical arbitrage strategy, concluding that this strategy is profitable in emerging markets, although this is not the case in emerging markets, demonstrating the weak form of efficiency. Balladares et al. (2021) extend the study by Sánchez-Granero et al. (2020) by taking stock prices of the 50 largest capitalization companies from 39 countries, demonstrating that the strategy works in emerging markets and when periods of high volatility occur. In addition, they analyse the pairs in terms of the value (H) of the Hurst exponent, given that the lower this value, the higher the return obtained. Currently, investors are valuing socially responsible financial investment positively, Montoya-Cruz et al. (2020) apply the Hurst Exponent selection method in the Pairs Trading strategy using FTSE4GOOD and FTS100 index securities, in this case, there are no significant differences between socially responsible and non-socially responsible portfolios. Ramos-Requena et al. (2021) study the Pairs trading strategy by selecting pairs based on stocks with a low degree of co-movement. They conclude that positive results can be achieved in bull markets and periods of low uncertainty.

16.3.2 HP Approach

As previously discussed, one of the distinguishing features of the statistical arbitrage strategy is the strong correlation between the pairs of securities comprising the portfolio. As a result, many studies employ correlation or cointegration methods when selecting pairs for the portfolio.

Ramos-Requena et al. (2020a) introduced a novel approach called HP, which aims to assess the co-movement between two time series. They utilize the Hurst exponent of the product series to gauge the degree of correlation between pairs of securities. Their study demonstrates that this innovative method is more effective at identifying relationships between assets compared to traditional correlation methods. The results indicate that HP outperforms the traditional correlation method in detecting asset relationships.

To begin, we establish the logarithmic return of a stock as follows:

$$r_n = \log\left(\frac{p_n}{p_{n-1}}\right)$$

where:

- p_n is the price of the asset in period n and
- p_{n-1} the price of the same asset in the previous period ($n - 1$)

Alternatively, we could calculate the series product as:

$$rs_n = \sum_{i=1}^n (r_i - \bar{r})(s_i - \bar{s})$$

where:

- $\bar{r}$ is the mean of the serie r_i .
- $\bar{s}$ is the mean of the serie s_i .

In the case of daily yields, there is no difference, as the averages are very close to 0.

We could calculate the degree of correlation between two stocks through the Hurst exponent H of the product series rs_n , depending on the value of H , it will tell us the following:

- If the value H is close to 0.5 the assets will have a low correlation.
- If the value H is close to 1 will mean that the assets have a high correlation.

We will call HP the Hurst exponent of the product series.

Since in a Pair Trading strategy we are looking for the two assets to have a high degree of co-movement, we can look for a pair of assets with a high H value (close to 1).

16.4 Methodology

The method for developing the statistical arbitrage strategy would consist of two parts. First, we would select the pairs using one of the methods defined in Sect. 16.2. Firstly, the share prices must be normalised, for which the methods defined in Ramos-Requena et al. (2020a, 2020b) can be used, which we will define below. Once normalised, we will apply the pair selection method, in most studies this phase usually takes 250 days. Once the pairs are selected, we will then apply the trading strategy, as explained in this section, this phase usually lasts about 120 days.

In the following, we will describe the different ways of calculating the b factor as defined by (Ramos-Requena et al., 2020b):

- Equal Weight (EW): Gatev et al. (2006) defined this method in their work on Pairs Trading, making it the most common procedure used to date. It consists of assuming that $b = 1$, i.e. the same amount is going to be invested in share A as in share B .

- Volatility: Developed by Ramos-Requena et al. (2017) is based on the idea that both components have the same volatility. Thus, if the volatility of r_A is $std(r_A)$ and the volatility of r_B is $std(r_B)$, if we want r_A and br_B to have the same volatility, then $b = std(r_A)/std(r_B)$.
- Based on the minimum distance of the logarithms of the prices: In this case we propose to minimise the function $f(b) = \sum_t |p_A(t) - bp_B(t)|$ to find the scaling factor b , such that p_A and bp_B have the minimum distance.
- Based on the correlation of returns: This method is based on the assumption that if the returns are correlated, then r_A is approximately equal to br_B , where b is obtained through the linear regression $r_A = br_B$.
- Based on price cointegration: It is based on the assumption that if prices (presented in logarithms) are cointegrated, then p_A is approximately equal to bp_B . Then b would be calculated through the linear regression $p_A = bp_B$.
- Based on the smallest Hurst exponent of the pair: Since the pair series is defined as $s(b)(t) = p_A(t) - bp_B(t)$, this methodology is based on finding the scaling factor b , such that the pair series $s(b)$ has the smallest Hurst exponent. This implies that this series will be as antipersistent as possible. In this way we will look for a value of b that minimises the function $f(b) = H(s(b))$ where $H(s(b))$ is the Hurst exponent for the $s(b)$ pair series.

Once we have selected the pair, we will describe the trading strategy we will carry out.

If s is the series of the pair, m is the moving average of s and σ is the standard deviation of $m - s$ then:

- When $s > m + \sigma$: we sell the pair. We close the position when $s < m$ or $s > m + n \cdot \sigma$.
- When $s < m - \sigma$: we buy the pair. Close the position when $s > m$ or $s < m - n \cdot \sigma$.

The value of n that multiplies sigma can vary, as shown by Fernández-Pérez et al. (2020), presenting a comparison of how results may vary depending on different values of n .

16.5 Conclusions

The area of econophysics is continuously developing. One of the noteworthy advancements in finance emerging from this field is the notion of the fractal market, challenging the conventional efficient market theory.

Financial markets deviate considerably from the behavior implied by the efficient market hypothesis, causing its forecasts to diverge from actual occurrences. Consequently, it is suggested that traditional analytical methods need adjustments to refine the results of predictive models. According to the fractal market hypothesis, prices demonstrate fractal characteristics, advocating for the application of fractal geometry in a more realistic approach to market analysis.

In this chapter, we have studied the application of this field in the realm of statistical arbitrage strategy. To do so, the approach based on the Hurst exponent has been employed. Among the various methodologies, the Generalized Hurst Exponent (GHE) has been used, given its simplicity of application. Additionally, the Hurst exponent has been applied to measure the degree of correlation of financial series products.

As demonstrated by Ramos-Requena et al. (2017, 2020a), this methodology is vastly superior and yields better results, measured through cumulative logarithmic returns, Sharpe ratio, Maximum Drawdown, compared to the different approaches used to date.

Furthermore, we have verified that employing the Hurst exponent method can serve to calculate the b factor that enables the normalization of the logarithmic price of stocks, as shown by Ramos-Requena et al. (2020b). It is a valuable tool, especially when the selection method known as the Hurst exponent is used.

Statistical arbitrage strategy is a relatively recent methodology and is highly attractive in the scientific domain. Its significant evolution over recent years has been observed in various domains such as asset markets, options and futures markets, cryptocurrency, and forex markets, etc.

We believe that continuing to apply different methodologies from the field of econophysics could render it more attractive in the future. Even current methods, such as cointegration and copulas, could be implemented using the Hurst exponent for their application.

The distance method outlined in this Sect. 16.2.1 has been employed thus far. Additionally, various methodologies designed for time series clustering may be applicable, such as the distance between ARIMA procedures (Piccolo, 1990), the GARCH distance (Otranto, 2008; Caiado and Crato, 2010), and more recently, those based on the Hurst exponent (Lahmiri, 2016; Cerqueti and Mattera, 2023), which may be relevant to the statistical arbitrage strategy.

Moreover, within the statistical arbitrage strategy, the well-known Pairs Trading strategy could be applied not only to two pairs of values but to three or more pairs of values to verify if applying this methodology remains profitable.

References

- Alexander, C., Giblin, I., & Weddington, W. (2002a). Cointegration and asset allocation: A new active hedge fund strategy. *Research in International Business and Finance*, 16(5), 65–90.
- Alexander, C., Giblin, I., & Weddington, W. (2002b). Cointegration and asset allocation: A new active hedge fund strategy. *Research in International Business and Finance*, 16(5), 65–90.
- Andrade, S., Di Pietro, V., & Seasholes, M. (2005). Understanding the profitability of pairs trading. Unpublished working paper, UC Berkeley, Northwestern University.
- Avellaneda, M., & Lee, J. H. (2010). Statistical arbitrage in the US equities market. *Quantitative Finance*, 10(7), 761–782.
- Balladares, K., Ramos-Requena, J. P., Trinidad-Segovia, J. E., & Sánchez-Granero, M. A. (2021). Statistical arbitrage in emerging markets: A global test of efficiency. *Mathematics*, 9(2), 179.

- Barabási, A. L., & Vicsek, T. (1991). Multifractality of self-affine fractals. *Physical Review A*, 44(4), 2730.
- Barunik, J., & Kristoufek, L. (2010). On Hurst exponent estimation under heavy-tailed distributions. *Physica A: Statistical Mechanics and its Applications*, 389(18), 3844–3855.
- Bowen, D., Hutchinson, M. C., & O'Sullivan, N. (2010). High-frequency equity pairs trading: Transaction costs, speed of execution, and patterns in returns. *The Journal of Trading*, 5(3), 31–38.
- Brunetti, M., & De Luca, R. (2023). Pairs trading in the index options market. *Eurasian Economic Review*, 13(1), 145–173.
- Bui, Q., & Ślepaczuk, R. (2022). Applying Hurst Exponent in pair trading strategies on Nasdaq 100 index. *Physica A: Statistical Mechanics and its Applications*, 592, 126784.
- Burgess, A. N. (1999). Statistical arbitrage models of the FTSE 100. *Computational Finance*, 99, 297–312.
- Caiado, J., & Crato, N. (2010). Identifying common dynamic features in stock returns. *Quantitative Finance*, 10(7), 797–807.
- Caldeira, J., & Moura, G. V. (2013). Selection of a portfolio of pairs based on cointegration: A statistical arbitrage strategy. Available at SSRN 2196391.
- Cerqueti, R., & Mattera, R. (2023). Fuzzy clustering of time series with time-varying memory. *International Journal of Approximate Reasoning*, 153, 193–218.
- Chan, Y. K. C. (2023). Predicting cointegration using neural network and fundamental ratios in statistical arbitrage.
- Chen, H., Chen, S., Chen, Z., & Li, F. (2019). Empirical investigation of an equity pairs trading strategy. *Management Science*, 65(1), 370–389.
- Clegg, M., & Krauss, C. (2018). Pairs trading with partial cointegration. *Quantitative Finance*, 18(1), 121–138.
- Conniffe, D., & Spencer, J. E. (2000). Approximating the distribution of the R/S statistic. *Economic and Social Review*, 31(3), 237–248.
- Da Silva, F. A. S., Ziegelmann, F. A., & Caldeira, J. F. (2018). Pairs trading: Optimizing via mixed copula versus distance method for S&P 500 assets.
- De Moura, C. E., Pizzinga, A., & Zubelli, J. (2016). A pairs trading strategy based on linear state space models and the Kalman filter. *Quantitative Finance*, 16(10), 1559–1573.
- Delgado, O. Y. Q., & Delgado, J. R. (2011). Estimación del exponente de Hurst y la dimensión fractal de una superficie topográfica a través de la extracción de perfiles. *UD y la geomática*, 5, 84–91.
- Di Matteo, T., Aste, T., & Dacorogna, M. M. (2005). Long-term memories of developed and emerging markets: Using the scaling analysis to characterize their stage of development. *Journal of Banking & Finance*, 29(4), 827–851.
- Diebold, F. X., & Rudebusch, G. D. (1989). Long memory and persistence in aggregate output. *Journal of Monetary Economics*, 24(2), 189–209.
- Dixon, M., Klabjan, D., & Bang, J. H. (2015). Implementing deep neural networks for financial market prediction on the Intel Xeon Phi. In *Proceedings of the 8th Workshop on High Performance Computational Finance* (pp. 1–6).
- Do, B., & Faff, R. (2010). Does simple pairs trading still work? *Financial Analysts Journal*, 66(4), 83–95.
- Do, B., & Faff, R. (2012). Are pairs trading profits robust to trading costs? *Journal of Financial Research*, 35(2), 261–287.
- Do, B., Faff, R., & Hamza, K. (2006). A new approach to modeling and estimation for pairs trading. In *Proceedings of 2006 Financial Management Association European Conference* (Vol. 1, pp. 87–99).
- Elliott, R. J., & Bradrania, R. (2018). Estimating a regime switching pairs trading model. *Quantitative Finance*, 18(5), 877–883.
- Elliott, R. J., Van Der Hoek, J., & Malcolm, W. P. (2005). Pairs trading. *Quantitative Finance*, 5(3), 271–276.

- Engelberg, J., Gao, P., & Jagannathan, R. (2009). An anatomy of pairs trading: The role of idiosyncratic news, common information and liquidity. In *Third Singapore International Conference on Finance*.
- Engle, R. F., & Granger, C. W. (1987). Co-integration and error correction: Representation, estimation, and testing. *Econometrica: Journal of the Econometric Society*, 55, 251–276.
- Fernández-Pérez, A., de las Nieves López-García, M., & Ramos-Requena, J. P. (2020). On the sensibility of the Pairs Trading strategy: The case of the FTS stock market index. *Studies of Applied Economics*, 38(3), 14.
- Ferreira, L. (2008). New tools for spread trading. *Futures*, 37(12), 38–41.
- Fischer, T., & Krauss, C. (2018). Deep learning with long short-term memory networks for financial market predictions. *European Journal of Operational Research*, 270(2), 654–669.
- Fischer, T. G., Krauss, C., & Deinert, A. (2019). Statistical arbitrage in cryptocurrency markets. *Journal of Risk and Financial Management*, 12(1), 31.
- Gatev, E., Goetzmann, W. N., & Rouwenhorst, K. G. (2006). Pairs trading: Performance of a relative-value arbitrage rule. *The Review of Financial Studies*, 19(3), 797–827.
- Guidolin, M., Hyde, S., McMillan, D., & Ono, S. (2009). Non-linear predictability in stock and bond returns: When and where is it exploitable? *International Journal of Forecasting*, 25(2), 373–399.
- Han, C., He, Z., & Toh, A. J. W. (2023). Pairs trading via unsupervised learning. *European Journal of Operational Research*, 307(2), 929–947.
- Hashemi, S. S., Hajiagha, S. H. R., Zavadskas, E. K., & Mahdiraji, H. A. (2016). Multicriteria group decision making with ELECTRE III method based on interval-valued intuitionistic fuzzy information. *Applied Mathematical Modelling*, 40(2), 1554–1564.
- Hassler, U. (1994). (Mis) specification of long memory in seasonal time series. *Journal of Time Series Analysis*, 15(1), 19–30.
- Hassler, U., & Wolters, J. (1995). Long memory in inflation rates: International evidence. *Journal of Business & Economic Statistics*, 13(1), 37–45.
- Hauke, J., & Kossowski, T. (2011). Comparison of values of Pearson's and Spearman's correlation coefficients on the same sets of data. *Quaestiones Geographicae*, 30(2), 87–93.
- Hendry, D. F., & Juselius, K. (2001). Explaining cointegration analysis: Part II. *The Energy Journal*, 22(1), 75–120.
- Hu, L. (2006). Dependence patterns across financial markets: A mixed copula approach. *Applied Financial Economics*, 16(10), 717–729.
- Huck, N. (2009). Pairs selection and outranking: An application to the S&P 100 index. *European Journal of Operational Research*, 196(2), 819–825.
- Huck, N. (2010). Pairs trading and outranking: The multi-step-ahead forecasting case. *European Journal of Operational Research*, 207(3), 1702–1716.
- Hurst, H. E. (1951). Long-term storage capacity of reservoirs. *Transactions of the American Society of Civil Engineers*, 116(1), 770–799.
- Jacobs, H., & Weber, M. (2015). On the determinants of pairs trading profitability. *Journal of Financial Markets*, 23, 75–97.
- Keshavarz Haddad, G., & Talebi, H. (2023). The profitability of pair trading strategy in stock markets: Evidence from Toronto stock exchange. *International Journal of Finance & Economics*, 28(1), 193–207.
- Khandani, A. E., & Lo, A. W. (2011). What happened to the quants in August 2007? Evidence from factors and transactions data. *Journal of Financial Markets*, 14(1), 1–46.
- Kim, S. H., Park, D. Y., & Lee, K. H. (2022). Hybrid deep reinforcement learning for pairs trading. *Applied Sciences*, 12(3), 944.
- Kleinow, T. (2002). Testing continuous time models in financial markets. Humboldt University.
- Krauss, C., Do, X. A., & Huck, N. (2017). Deep neural networks, gradient-boosted trees, random forests: Statistical arbitrage on the S&P 500. *European Journal of Operational Research*, 259(2), 689–702.
- Krauss, C., & Stübinger, J. (2017). Non-linear dependence modelling with bivariate copulas: Statistical arbitrage pairs trading on the S&P 100. *Applied Economics*, 49(52), 5352–5369.

- Lahmiri, S. (2016). Clustering of Casablanca stock market based on hurst exponent estimates. *Physica A: Statistical Mechanics and its Applications*, 456, 310–318.
- Liew, R. Q., & Wu, Y. (2013). Pairs trading: A copula approach. *Journal of Derivatives & Hedge Funds*, 19, 12–30.
- Lin, B., & Tan, Z. (2023). Exploring arbitrage opportunities between China's carbon markets based on statistical arbitrage pairs trading strategy. *Environmental Impact Assessment Review*, 99, 107041.
- Litterman, B. (2004). *Modern investment management: An equilibrium approach*. John Wiley & Sons.
- Lo, A. W., & MacKinlay, A. C. (1988). Stock market prices do not follow random walks: Evidence from a simple specification test. *The Review of Financial Studies*, 1(1), 41–66.
- Lo, A. W., & MacKinlay, A. C. (1997). Maximizing predictability in the stock and bond markets. *Macroeconomic Dynamics*, 1(1), 102–134.
- López-García, M. N., & Ramos-Requena, J. P. (2019). Different methodologies and uses of the Hurst exponent in econophysics. *Studies of Applied Economics*, 37(2), 96–108.
- Lu, J. Y., Lai, H. C., Shih, W. Y., Chen, Y. F., Huang, S. H., Chang, H. H., & Dai, T. S. (2022). Structural break-aware pairs trading strategy using deep reinforcement learning. *The Journal of Supercomputing*, 78(3), 3843–3882.
- Mandelbrot, B. (1972). Statistical methodology for nonperiodic cycles: From the covariance to R/S analysis. *Annals of Economic and Social Measurement*, 1(3), 259–290.
- Mandelbrot, B. B. (1971). When can price be arbitrated efficiently? A limit to the validity of the random walk and martingale models. *The Review of Economics and Statistics*, 53, 225–236.
- Mandelbrot, B. B., & Mandelbrot, B. B. (1997). Discontinuity and scaling: Their scope and likely limitations. *Fractals and Scaling in Finance: Discontinuity, Concentration, Risk. Selecta Volume E*, 50–78.
- Mandelbrot, B. B., & Wallis, J. R. (1969). Some long-run properties of geophysical records. *Water Resources Research*, 5(2), 321–340.
- Montoya-Cruz, E., Ramos-Requena, J. P., Trinidad-Segovia, J. E., & Sánchez-Granero, M. Á. (2020). Exploring arbitrage strategies in corporate social responsibility companies. *Sustainability*, 12(16), 6293.
- Moritz, B., & Zimmermann, T. (2014). Deep conditional portfolio sorts: The relation between past and future stock returns. LMU Munich and Harvard University Working paper.
- Movahed, M. S., & Hermanis, E. (2008). Fractal analysis of river flow fluctuations. *Physica A: Statistical Mechanics and its Applications*, 387(4), 915–932.
- Mrčela, L., Mercep, A., Begusic, S., & Kostanjcar, Z. (2017). Portfolio optimization using preference relation based on statistical arbitrage. In *2017 International Conference on Smart Systems and Technologies (SST)* (pp. 161–165).
- Otranto, E. (2008). Clustering heteroskedastic time series by model-based procedures. *Computational Statistics & Data Analysis*, 52(10), 4685–4698.
- Peng, C. K., Buldyrev, S. V., Goldberger, A. L., Havlin, S., Sciortino, F., Simons, M., & Stanley, H. E. (1992). Long-range correlations in nucleotide sequences. *Nature*, 356(6365), 168–170.
- Perlin, M. S. (2009). Evaluation of pairs-trading strategy at the Brazilian financial market. *Journal of Derivatives & Hedge Funds*, 15, 122–136.
- Piccolo, D. (1990). A distance measure for classifying ARIMA models. *Journal of Time Series Analysis*, 11(2), 153–164.
- Poterba, J. M., & Summers, L. H. (1988). Mean reversion in stock prices: Evidence and implications. *Journal of Financial Economics*, 22(1), 27–59.
- Preston, T. (2005). Pairs trading. Obtained online from www.trader-mag.com. Cited 15 Nov 2023.
- Ramos-Requena, J. P., López-García, M. N., Sánchez-Granero, M. A., & Trinidad-Segovia, J. E. (2021). A cooperative dynamic approach to pairs trading. *Complexity*, 2021, 1–8.
- Ramos-Requena, J. P., Trinidad-Segovia, J. E., & Sánchez-Granero, M. A. (2017). Introducing Hurst exponent in pair trading. *Physica A: Statistical Mechanics and its Applications*, 488, 39–45.

- Ramos-Requena, J. P., Trinidad-Segovia, J. E., & Sánchez-Granero, M. Á. (2020a). An alternative approach to measure co-movement between two time series. *Mathematics*, 8(2), 261.
- Ramos-Requena, J. P., Trinidad-Segovia, J. E., & Sánchez-Granero, M. Á. (2020b). Some notes on the formation of a pair in pairs trading. *Mathematics*, 8(3), 348.
- Sánchez, M. Á., Trinidad, J. E., García, J., & Fernández, M. (2015). The effect of the underlying distribution in Hurst exponent estimation. *PLoS One*, 10(5), e0127824.
- Sánchez-Granero, M. A., Balladares, K. A., Ramos-Requena, J. P., & Trinidad-Segovia, J. E. (2020). Testing the efficient market hypothesis in Latin American stock markets. *Physica A: Statistical Mechanics and its Applications*, 540, 123082.
- Sánchez-Granero, M. J., Fernández-Martínez, M., & Trinidad-Segovia, J. E. (2012). Introducing fractal dimension algorithms to calculate the Hurst exponent of financial time series. *The European Physical Journal B*, 85, 1–13.
- Šarić, F., Begušić, S., Merćep, A., & Kostanjčar, Z. (2023). Statistical arbitrage portfolio construction based on preference relations. *Expert Systems with Applications*, 238, 121906.
- Stander, Y., Marais, D., & Botha, I. (2013). Trading strategies with copulas. *Journal of Economic and Financial Sciences*, 6(1), 83–107.
- Stanley, H. E., Amaral, L. A. N., Canning, D., Gopikrishnan, P., Lee, Y., & Liu, Y. (1999). Econophysics: Can physicists contribute to the science of economics?. *Physica A: Statistical Mechanics and its Applications*, 269(1), 156–169.
- Stübinger, J., & Bredthauer, J. (2017). Statistical arbitrage pairs trading with high-frequency data. *International Journal of Economics and Financial Issues*, 7(4), 650–662.
- Stübinger, J., Mangold, B., & Krauss, C. (2018). Statistical arbitrage with vine copulas. *Quantitative Finance*, 18(11), 1831–1849.
- Tadi, M., Abkar, M., & Motaharinia, V. (2018). Evaluation of pairs trading strategy using distance approach at Tehran stock exchange. *Journal of Investment Knowledge*, 7(26), 99–112.
- Takeuchi, L., & Lee, Y. Y. A. (2013). Applying deep learning to enhance momentum trading strategies in stocks. Technical Report. Stanford University, Stanford.
- Vidyamurthy, G. (2004). *Pairs trading: Quantitative methods and analysis*. John Wiley & Sons.
- Wong, B. (2010). Trading trading strategies: Pairs trading. Singapore Stock Exchange.
- Xie, W., Liew, R. Q., Wu, Y., & Zou, X. (2015). Pairs trading with copulas. SSRN.
- Zebedee, A. A., & Kasch-Haroutounian, M. (2009). A closer look at co-movements among stock returns. *Journal of Economics and Business*, 61(4), 279–294.
- Zhao, Z., Zhou, R., & Palomar, D. P. (2019). Optimal mean-reverting portfolio with leverage constraint for statistical arbitrage in finance. *IEEE Transactions on Signal Processing*, 67(7), 1681–1695.
- Zhao, Z., Zhou, R., Wang, Z., & Palomar, D. P. (2018). Optimal portfolio design for statistical arbitrage in finance. In *2018 IEEE Statistical Signal Processing Workshop (SSP)* (pp. 801–805).