QFE, 9(4): 912–932.
DOI:10.3934/QFE.2025032
Received: 11 June 2025
Revised: 05 September 2025
Accepted: 25 November 2025
Published: 09 December 2025

Quantitative Finance and Economics

https://www.aimspress.com/journal/QFE

Research article

Exploration of multiple asset investment opportunities based on pair trading

Mariano González-Sánchez[1,*], **Juan M. Nave Pineda**[2] **and Javier Ruiz Rincón**[2]

[1] Department of Business and Accounting, National Distance Education University (UNED), Paseo Senda del Rey, 11, 28040 Madrid, Spain

[2] Economic Analysis and Finance Department, Castilla La Mancha University (UCLM), Avenida de los Alfares, 42, 16071 Cuenca, Spain

* **Correspondence:** Email: mariano.gonzalez@cee.uned.es; Tel: +34913986366; Fax: +34917705365.

Abstract: The financial literature on pair trading shows that Johansen's methodology is the most consistent for finding long-term cointegrated assets. Previous work has either looked for cointegrating relationships of total assets when it was feasible to identify them or has only analyzed pairwise relationships. We present a bottom-up analysis of the possible cointegrating relationships among assets to search for the most profitable strategies. On monthly prices of stocks (Eurostosxx-50), we find that cointegration relationships go beyond two assets, have volatile behavior, and different terms. Both short-term and long-term investments show returns higher than the benchmark index.

Keywords: pair-trading; cointegration; mean reversion; long-run relationship; statistical arbitrage

JEL Codes: C32, C58, G11, G12

1. Introduction

Pair trading is a strategy that seeks to take advantage of changes in the relationship that exists between pairs of assets, so that if there is a significant change in the relationship, a position would be taken under the premise that this relationship is stable over the long term, and therefore, a reversal to the mean level will occur (see Gatev et al. (2006)). Justifications for pair trading are diverse for example, Chen et al. (2019) found that the pairs trading profits are explained by the short-term reversal and industry momentum but, most of the literature analyzing pair trading justifies it due to the possibility of arbitrage. According to Bondarenko (2003), arbitrage exists when a self-funded portfolio has no potential losses, i.e., a pure arbitrage portfolio. However, Hogan et al. (2004) formulate statistical arbitrage for an initial zero-cost, self-funded, positive expected present value portfolio and, with sample variance tending to zero if the probability of loss is non-zero, for a finite specific moment in time, therefore, it is zero beta or outside systematic risk. The practical application of this pairwise strategy can basically be grouped into three typologies (Krauss (2017) is a review): studies of the distance between the normalized prices of the assets; analysis of the linear or non-linear dependence between the returns of assets; and one looks for the long-term equilibrium relationships between the asset prices:

- The distance methodology consists of investing or disinvesting in the asset pair depending on how the value of the spread between the normalized prices deviates by a certain amount, measured in terms of standard deviation (for example, Elliott et al. (2005), Gatev et al. (2006), and Bowen et al. (2010)).
- The dependence analysis methodology moves from simple linear correlation (see Stander et al. (2013)) to more complex models that adjust to the outliers of the daily returns using copulas (for example, Krauss and Stübinger (2017)). This trading strategy takes advantage of the relative mispricing dependence between a pair of stocks and involves taking a position on the stocks when they diverge from their historical relationship[1].
- The cointegration technique analyzes the equilibrium models among prices, i.e., long-term equilibrium relationships are sought among assets, so that although prices are not stationary or I(1), this relationship is stationary and, also acts as a corrective (error correction model) on the behavior of individual prices, i.e., this strategy tries to take advantage of the common trends observed among prices and, when they move away from it, to take positions waiting for a reversion to the mean trend (for further details see Vidyamurthy (2004)).

Note that all methodologies attempt to take advantage of situations in which prices move away from the mean trend, but while the distance and dependence methodologies work on stationary series (returns), in which long-term trends and reversion are not guaranteed, cointegration analyzes non-stationary series (prices) and aims to find those long-term trends. In addition, there are studies that find that the results obtained by applying dependence analysis (copulas) and cointegration are similar, while the distance methodology seems to present the worst results (see Rad et al. (2016) and Haddad and Talebi (2023)).

There are two commonly used cointegration tests: Engle and Granger (1987) and Johansen (1991). However, Engle and Granger (1987) is only applicable to pairs of assets (see Bowen et al. (2010) and Hanson and Hall (2012)), and is not useful for multivariate analysis and this test is sensitive to non-normality and other statistical characteristics of the time series (see Stock (1987), Stock and Watson (1993) and Lee and Hur (2021)); Johansen (1991), however, is applicable to multivariate analysis and the results are consistent against non-normal errors and the ARCH/GARCH process (there are numerous studies comparing the two techniques with superior results for Johansen, see for example, Bilgili (1998), Silvapulle and Podivinsky (2000), Cavaliere et al. (2010), Kurita (2013), Chan (2013)).

The empirical studies using Johansen's methodology to analyze long-run equilibrium relationships between financial variables can be grouped into two cases depending on whether or not the relationship is identifiable (whether or not there is an a priori expected value for the parameters of the linear combination of cointegrated assets):

[1] Other proposers use the exponent Hurst (see Ramos et al. (2017); however, the persistence of two series does not guarantee the existence of cointegration between them; that is, that they are jointly predictable (stationary) while individually they are not (non-stationary).

- If the parameters of the cointegration relationship are identified based on some theory or model, for example, triangular arbitrage by cross-exchange currency (see Ferré and Hall (2002)), since the parameters are 1 in absolute value.
- Otherwise, the parameters are not known a priori and, their expected values cannot be determined. In these cases, we observe, in the financial literature, two approaches to the problem: either to analyze the cointegration relationships between the whole set of variables or assets (for example, Perlin (2007) tries to study it but does not apply the Johasen methodology, only performs a multiple regression): Sheng and Tu (2000) analyzed 12 index stock markets, Östermark (2001) studied 6 variables and Kim (2003) analyzed 5 variables; or to study the cointegration relationships by pairs of variables (Vidyamurthy (2004)).

Obviously, the pair trading strategy is included in the latter case, since the expected value of the parameters of the cointegration relationship cannot be determined a priori. Then, when we opt to analyze the pairwise cointegration relationships (bottom approach), although, the relationship is not initially identified, if there is only one possible cointegration relationship, it is unique, since the maximum number of existing relationships is the total assets minus 1; therefore, the advantage of the bottom approach is that the relationship is unique and the identification of the parameters can be tested more easily. The disadvantage is that higher-order relationships (more than two assets), which could be more profitable, are no longer studied. Analyzing relationships that include all assets (up approach) has the disadvantage that the identification of the parameters is practically impossible if the relationship is not unique (only one)[2].

In this context, we propose a bottom-up methodology to narrow down the set of possible cointegration relationships, starting with those of lower rank (pair trading) and gradually increasing the number of assets participating in the cointegrated portfolio. Ramos Jungblut (2024) performs a similar analysis but arbitrarily limits the maximum number of co-integrated assets to ten, and the assets were randomly selected. Then our aim is to go beyond the pairwise relationship to not leave out an infinite number of possible combinations of more than two assets that, being cointegrated, could be more profitable than simple pair trading and closer to the statistical arbitrage portfolio. Since the possible asset combinations grow exponentially with increasing numbers of assets considered, and since if the relationship is not unique it is almost impossible to identify the relationship, we propose a methodology that not only identifies unique cointegrating relationships of increasing order but also reduces the computational effort, allowing us to explore a much larger set of possible asset combinations than simple pair trading or linear combinations of total assets.

The rest of the paper is organized as follows: Section 2 explains the methodology applied in the study and describes the sample; Section 3 shows the main results and, finally, concludes with the main findings.

[2]Note that,when more assets are involved, the combinations grow exponentially. For example, three assets could have three pair cointegrating relationships and one cointegration relationship for all three at once. Additionally, the transitive property is not characteristic of cointegrating relationships, i.e., if A and B are cointegrated and B-C are also cointegrated, that does not mean that A-C are cointegrated, and therefore it also does not mean that A-B-C are cointegrated in one relationship at a time (Ferré (2004) showed that this depended on the proportionality between the residual variances of the Vector Autoregressive or VAR).

2. Materials and method

2.1. Methodology

Suppose that Y_t is a vector of prices of M assets at instant t ($t = 1, ..., T$), so that using the usual nomenclature of a Vector Error Correction Model (VECM) would result in:

$$\Delta Y_t = \mu + \sum_{s=1}^{S} \Gamma_s \cdot \Delta Y_{t-s} + \Pi \cdot Y_{t-1} + u_t \tag{1}$$

where ΔY_t is the first difference in prices, μ is a vector of M-constants, s represents s-lag, S is maximum lags estimated by an information criterion (e.g. AIC), Γ_s is a vector of parameters which captures the effects of the s-lagged dependent variable and $\Pi = \alpha \cdot \beta^{'}$ is the matrix that collects the error correction effect, being the α matrix the effect on each asset of the long-run equilibrium relation or adjustment coefficients, while the β matrix contains the weights of the lagged prices of each asset in this relation (or cointegrating vectors). To this expression, we apply Johansen's procedure (Johansen (1991)), in particular, the trace test. This test allows us to determine the cointegrating relationships that exist ($r = 0, ..., M - 1$); thus, we test the null hypothesis of $r \leq k$ versus $r > k$. And, u_t is an error vector.

Our objective is to find the cointegrating relationships of N assets ($N = 2, ..., M$) are unique. This means that the hypothesis (H_0) $r \leq 0$ vs. $r > 0$ is rejected and that (H_1) $r = 1$ vs. $r > 1$ is not rejected, since this allows us to easily check that the weights of the cointegrating vector are non-zero. To do so, we require a *p-value* ≤ 0.01 (estimated according to Doornik (1998)) to reject H_0 and a *p-value* ≥ 0.1 not to reject H_1, and then, we test $\beta_i \neq 0$ by a log-likelihood ratio test (see expression 5.11 Johansen (1991)).

Intuitively, Johansen's procedure consists of finding the non-zero eigenvalues of $\Pi^{'} \cdot \Pi$ (used to estimate the trace test), so that their corresponding normalized eigenvectors are the weights of the asset prices in the cointegrating relationship. Although there could be more cointegrating relationships of N assets, our requirement of only one is for various reasons: to maintain the same criterion from the 2-asset to the N-asset portfolios; at the same time, to reduce the computational effort since the cointegrating relationships are not defined and therefore any rotation of the parameters would also be another cointegrated portfolio with the same assets; and when the relationship is unique the identification of the relationship (parameter values) is simpler, and also, to test that they are statistically significant, since the dimension of the problem is reduced to testing a hypothesis. Note that this joint hypothesis for a portfolio with N assets involves testing whether $\forall i \neq k \quad \theta_i = 0$, where k-asset is the one in which we normalize the eigenvectors ($\theta_k = 1$)[3]. This log-likelihood ratio test is distributed as a χ^2 with $N - 1$ degrees of freedom.

The main problem is that as the number of assets (M) involved in cointegration relationships grows, the combinations increase exponentially. So, it is a problem of combinations of M elements taken from N in N without replacement; in particular, the possible combinations for each case can be estimated as $\frac{M!}{N! \cdot (M-N)!}$[4]. To understand the methodological proposal, let us assume that a set of M assets can

[3]If θ_i are the eigenvectors associated with the non-zero eigenvalue of the cointegrating relationship found, then we normalize so that the initial value of the cointegrated portfolio is as close as possible to zero to obtain a self-funded portfolio according to the statistical arbitrage conditions. Thus, if the cheapest combination (cost close to zero) is obtained by normalizing on the k-asset, then the normalized weights are $\theta_i = \frac{\beta_i}{\beta_k}$.

[4]Note, for example, that if $M = 10$ assets are available, the possible cointegrating relationships can be $N = 2, .., 10$ assets, so that there are 1,013 possible cointegrating relationships to be analyzed.

only have $M - 1$ cointegrating relationships, but if we analyze them in sets of a lower order (2 to 2, 3 to 3, ...), then the total number of cointegrating relationships could exceed this $M - 1$ value[5]. Therefore, if we analyze the total, we may be leaving out an important set of cointegration relationships from the investment opportunities. So, our proposal starts with the lowest possible order (pairs) and increases gradually. Furthermore, the assumption to be made is that the relationships of N assets are a linear combination of the pairwise relationships. Based on this premise, as we increase the size of the possible cointegrated portfolio, we exclude those assets from the analysis that have not participated in cointegration relationships of a lower order and then considerably reduce the computational effort. At the same time, it is observed at each time instant of analysis, that some assets are involved in a greater number of cointegration relationships, so that when the size of the possible cointegrated portfolio is increased, we begin by analyzing the combinations that include these assets (pivot assets), thereby also achieving a considerable reduction in computation time. To make this problem computationally tractable, we operate as follows:

1. We analyze the possible pairwise cointegration relationships. In this way we obtain not only the cointegrated pair trading, but we also observe which assets are not cointegrated with any other asset, and by eliminating them from the total assets (M) we obtain a subsample M_2 of pairwise cointegrated assets with a unique relationship and, we order the assets according to the number of cointegration relationships in which they participate (highest to lowest).
2. Afterwards, we then perform the N-assets ($N = 3, ..., M$) cointegration analysis within the M_{N-1} subsample, and we construct an M_N subsample consisting only of the assets that participate in any of the unique N-assets cointegrating relationships and ordered according to the number of cointegrating relationships in which they participate (highest to lowest), as we did before. The procedure stops when a subset M_N is empty.

Finally, once the cointegration relationships between sets of N assets have been found, we test the hypothesis of non-null parameters and, we look for those that initially meet the statistical arbitrage conditions, i.e., self-funded ($\mu_r = 0$) and without systematic risk ($\beta_{r,market} = 0$); for this purpose we perform the following optimization among of all cointegration relationships at each time instant analyzed:

$$\min(\mu_r^2 + \beta_{r,market}^2) \tag{2}$$

where r is each of the total R cointegrating relationships found.

2.2. Data

The sample for our empirical study is composed of monthly quotes from companies that are part of the Eurostoxx-50 and whose prices are available from January 2002 to December 2024, inclusive. The data were obtained from the Thomson Reuter's Refinitiv database, and all prices are in Euros. Our study is conducted on a monthly basis for several reasons. Firstly, because of the higher consistency of the cointegration tests; also because the buy-to-hold strategy is the most profitable for this type of term as shown by Bağci and Soylu (2024), finally because it reduces the costs for an investor compared to

[5]For example, if we have three assets, there can be a maximum of two cointegrating relationships, while if we analyze them in pairs, we can find up to three relationships.

high frequency trading (see Tokat and Hayrullahoğlu (2022)). Additionally, daily data are typically subject to higher noise, which reduces the power of statistical tests and may result in either spurious rejection or spurious acceptance of cointegration, as noted in the references reviewed above. In addition, the increased incidence of structural breaks further exacerbates the problem, potentially yielding spurious findings of cointegrating relationships. To our knowledge, there is no consistent methodology to determine cointegration relationships under such extreme conditions for high-frequency data.

The data are not filtered to avoid complications and price jumps or structural breaks such as those observed during a financial crisis, since, as justified by Tokat and Hayrullahoğlu (2022), the aim of imitating a realistic trading environment under uncertainty.

In Table-1, we include the name of the companies, abbreviation and the ADF test (Augmented Dickey-Fuller test) to check that all the data in levels (prices) are non-stationary, and, since the return (equivalent to taking the first difference of the logarithm) is known to be stationary, then it is evident that all the prices are $I(1)$ or integrated of order 1.

Table 1. Sample companies and stationarity test

Name	*Abrev.*	*ADF test*	*Name*	*Abrev.*	*ADF test*
Adidas	*ADIDAS*	-1.035	*Infineon Technologies AG*	*INFINEON*	-1.049
Ahold Delhaize	*AHOLD*	-0.500	*ING Banck*	*ING*	-2.148
Air Liquide	*AIRLIQ*	-0.644	*Intesa Sanpaolo*	*INTENSA*	-1.854
Airbus	*AIRBUS*	-0.627	*Kering*	*KERING*	-1.387
Allianz SE	*ALLIANZ*	-1.434	*L'Oreal SA*	*LOREAL*	-0.600
Anheuser – Busch Inbev SA	*ANHEUSER*	-1.588	*Mercedes – Benz Group AG*	*MERCEDES*	-2.541
ASML	*ASML*	0.468	*Moet Hennessy Louis Vuitton*	*LVHM*	-0.336
AXA SA	*AXA*	-2.340	*Muenchener Rueckversicherungs*	*MUENCHENER*	1.197
BASF	*BASF*	-2.179	*Nokia*	*NOKIA*	-1.839
Bayer AG	*BAYER*	-1.170	*Nordea Bank*	*NORDEA*	-2.453
BBVA Bank	*BBVA*	-1.859	*Pernod Ricard*	*RICARD*	-1.827
BMW	*BMW*	-2.185	*Safran*	*SAFRAN*	1.044
BNP Paribas SA	*BNP*	-2.239	*Sanofi*	*SANOFI*	-1.414
Danone	*DANONE*	-2.546	*Santander Bank*	*SANTANDER*	-2.247
Deutsche Boerse AG	*DEUSCBORSE*	-0.335	*SAP SE*	*SAP*	2.005
Deutsche Post AG	*DEUSCPOST*	-2.022	*Schneider Electric*	*SCHENEIDER*	1.342
Deutsche Telekom AG	*DEUSCTELEK*	-0.652	*Siemens AG*	*SIEMENS*	-2.689
Enel	*ENEL*	-2.290	*Stellantis*	*STELLANTIS*	-1.957
Eni	*ENI*	-1.689	*Total Energies*	*TOTENERG*	-2.545
Essilor Luxottica SA	*ESSILOR*	0.802	*UniCredit*	*UNICREDIT*	-2.009
Hermes International SCA	*HERMES*	1.958	*Vinci*	*VINCI*	-1.592
Iberdrola	*IBERDROLA*	-0.618	*Volkswagen AG*	*VOLKSWAGEN*	-2.114
Inditex	*INDITEX*	-0.171	*Eurostoxx* – 50	*EURO* – 50	-1.052

Note: *ADFtest* is the result for the best delay according to the AIC criterion and with constant. The critical value of the ADF test at 5% confidence level is -2.87.

Note that the maximum number of assets that can make up a co-integrated portfolio is 45, which to the best of our knowledge, is the highest number analyzed so far.

3. Results

3.1. Cointegration analysis

We start the estimate of the possible cointegration relationships in December 2014, that is, the first estimate is done with 156 monthly observations (sample start January 2002 and more than 100 data as stated in the literature reviewed above) or 13 years. The successive monthly estimates incorporate one more data. So that the last estimate (December 2024) is performed with the whole sample period, i.e., 276 monthly observations or 23 years. A cointegration analysis is performed monthly for 121 months (more than 10 years), from December 2014 to December 2024. We consider an increasing sample size to avoid the high sensitivity of the results to the sampling window (see Tokat and Hayrullahoğlu (2022) Table-3), and we avoid the problem of survival (or new entries) of assets.

In that period, we find 31,955 cointegrating relationships formed by two or more assets in the sample, i.e., the trace test rejects those where there are 0 cointegrating relationships and where there is more than 1 cointegrating relationship; therefore, only one relationship exists. If we perform the cointegration analysis of the 45 assets jointly, the trace test (for 121 months) results in between 22 and 43 cointegration relationships as minimum and maximum, respectively. Therefore, addressing the problem from a top-down approach would leave out an important set of cointegrated portfolios.

Table-2 shows summary statistics (percentiles) of the monthly results obtained from the analysis of the cointegration relationships found.

Table 2. Statistical analysis by percentiles of monthly cointegration relationships

Estimation per month	*Min*	25%	50%	75%	*Max*
number relations	5	32	56	127	5,308
Assets cointegrated	4	12	16	19	26
Assets NO cointegrated	19	26	29	33	41
Trace test (*relation* = 0)	19.622	82.828	137.716	190.582	530.219
pvalue (*relation* = 0)	0.000	0.000	0.000	0.003	0.010
Trace test (*relation* = 1)	0.010	27.756	62.141	106.064	279.172
pvalue (*relation* = 1)	0.107	0.152	0.250	0.454	0.892
mean value	53.958	18.131	153.754	726.539	1,718.829
std. dev.	97.208	30.247	152.714	504.526	796.268
Normal test	5.701	14.840	43.486	184.925	102.947
pvalue	0.500	0.000	0.000	0.000	0.000
Robust Ljung Box test in raw	0.397	0.468	1.610	3.260	106.359
pvalue	0.501	0.396	0.079	0.040	0.000
Robust Ljung Box test in square	0.328	4.233	6.031	10.012	16.079
pvalue	0.486	0.497	0.250	0.042	0.023
LM ARCH test (*Engle*)	1.697	2.054	2.364	3.102	36.803
pvalue	0.118	0.066	0.039	0.027	0.000
ADF test with constant	-6.159	-4.246	-3.066	-3.989	-8.010

Table-2 shows that the minimum number of cointegrating relationships found in a month is 5, while the maximum is 5,308 relationships. This already shows that computationally this type of strategy requires a significant amount of resources and time, so that its use in short time intervals is limited by the available resources. More importantly, though, is the evidence found that equilibrium relationships go through different moments in time: in some, there are few investment opportunities, while in others, there are numerous cointegration relationships. This means that the investment opportunities of the pair (or more) trading strategy are not constant over time. We also observe that despite the fact that the companies in the sample belong to a common economic area and have the same currency, there is a set of assets that is not permanently cointegrated with the rest (a minimum of 19 and a maximum of 41 in each month). This also points to the possibility that a cointegration relationship is not permanent over time, an issue that we analyze later. With respect to the trace test, we found that in all cases there is a one-to-one relationship, since we reject the null hypothesis of 0 cointegration relationships at a confidence level of 1% and we also rejected that the cointegration rank is 2 or more at a confidence level higher than 10%, in all cases.

The mean value and standard deviation of the cointegration relationships shows their average value and a high volatility with respect to this mean value. This means that the value of each cointegration relationship is very volatile around the mean (expected convergence value) and thus the proposed strategy of testing when it is below the mean by a large percentage and selling when it exceeds the mean by a significant amount can lead to significant profits.

Finally, the statistical tests show, as expected, that around half of the cointegrating relationships show autocorrelation in levels and the square of the values, but most of them do not follow a normal distribution, probably because of their high volatility, which leads to outliers, and this is evidenced by the fact that in some cases the relationship is rejected as passing the ARCH test at a 5% confidence level

Table-3 shows the minimum and maximum number of cointegration relationships found to be a function of the number of assets involved. It also indicates the number of times a relationship with that rank occurs during the 121 months of analysis, and it contains the assets forming the most frequent relationship.

Table 3. Monthly cointegration relationships according to the number of participating assets

Estimation per month	*Min*	*Max*	*Total months*	*Assets for most repeat cointegration*
Number relation with 2 *assets*	3	43	121	
Times in months	1	119		Ahold-Essilor
Number relation with 3 *assets*	8	162	119	
Times in months	1	60		Vinci-Muenchener-Ricard
Number relation with 4 *assets*	6	482	107	
Times in months	1	42		Vinci-Muenchener-Ricard-Airliq
Number relation with 5 *assets*	2	1102	97	
Times in months	1	24		Vinci-Muenchener-Ricard-Airliq-Loreal
Number relation with 6 *assets*	3	1886	79	
Times in months	1	12		Danone-Ricard-Muenchener-Essilor-Airliq-Loreal
Number relation with 7 *assets*	8	2392	59	
Times in months	1	11		Inditex-Ricard-Anheuser-Bayer-Essilor-Loreal
Number relation with 8 *assets*	3	2283	40	
Times in months	1	8		Inditex-Ricard-Anheuser-Bayer-Muenchener-Essilor-Vinci-Loreal
Number relation with 9 *assets*	2	1598	25	
Times in months	1	5		Basf-Inditex-Ricard-Anheuser-Bayer-Muenchener-Essilor-Vinci-Loreal
Number relation with 10 *assets*	2	894	16	
Times in months	1	4		Basf-Inditex-Ricard-Anheuser-Bayer-Muenchener-Essilor-Vinci-Loreal-Sap
Number relation with 11 *assets*	1	354	10	
Times in months	1	3		Sap-Airliq-Scheneider-Vinci-Essilor-Bayer-Anheuser-Muenchener-Basf-Ricard-Danone
Number relation with 12 *assets*	1	94	3	
Times in months	1	3		Sap-Loreal-Scheneider-Vinci-Essilor-Anheuser-Muenchener-Deuscpost-Basf-Ricard-Inditex-Danone
Number relation with 13 *assets*	1	17	2	
Times in months	1	2		Sap-Loreal-Scheneider-Vinci-Essilor-Anheuser-Muenchener-Deuscpost-Basf-Ricard-Inditex-Danone-Bayer
Number relation with 14 *assets*	1	1	1	
Times in months	1	1		Asml-Sap-Loreal-Airliq-Airbus-Vinci-Safran-Bayer-Anheuser-Enel-Inditex-Kering-Deuscborse-Unicredit

From the results in Table-3, we can see that 2-asset relationships occur every month (121), but 3-asset relationships occur on 119 occasions, and even 14-asset relationships occur on at least one occasion. Thus, limiting the strategy to pair trading (two assets) means excluding from the set of investment strategy opportunities an enormous volume of possibilities, with the consequent underestimation of potential profits. Therefore, our methodology, unlike the previous ones, includes in its analysis a wider range of investment options. We also note that some assets are more common than others in the cointegrating relationships, e.g., *Ricard*, *Muenchener*, and *Vinci* are in almost all maximum repeated cointegrated portfolios; therefore, we find that some assets show a stable stationary relationship and others an unstable one.

As a consequence of the above, Table-4 shows the most repeated assets in the cointegration relationships for each of the possible groupings (from 2 to 14 assets).

Table 4. Most repeated assets in the different cointegrated portfolios

Asset	Assets 2	Assets 3	Assets 4	Assets 5	Assets 6	Assets 7	Assets 8	Assets 9	Assets 10	Assets 11	Assets 12	Assets 13	Assets 14
ASML		*	*	*	*	*	*	*					*
LVHM						*							
TOTENERG							*						
SAP									*	*	*	*	*
SIEMENS													
LOREAL				*		*	*		*	*		*	*
AIRLIQ	*					*	*				*		*
SCHENEIDER			*							*	*	*	
AIRBUS							*	*					*
IBERDROLA						*							
VINCI					*		*	*	*	*	*	*	*
SAFRAN								*					*
ESSILOR								*	*	*	*	*	
BAYER				*	*	*		*	*		*	*	*
INFINEON													
ANHEUSER		*			*		*	*	*	*	*	*	*
MUENCHENER									*	*	*	*	
DEUSCPOST										*	*	*	
BASF			*	*					*	*	*	*	
RICARD									*	*	*	*	
INDITEX					*				*			*	*
STELLANTIS					*	*		*					
BMW								*					
DANONE										*	*	*	
DEUSCBORSE													*
AHOLD	*	*	*	*									
UNICREDIT							*						*

Table-4 shows the most repeated assets according to the size of the cointegration portfolio, e.g., in cointegrated portfolios of three assets, the three most common assets in these relationships appear, even though there is no cointegration relationship between them. These results are interesting, since we can see how, for example, *ASML* is an asset that often cointegrates with the rest of the assets for different sizes (from 3 to 9 assets), but instead, this asset is not part of the most repeated cointegrated portfolio for these sizes (see Table-3). Logically, for the ratio of 14 assets, the composition coincides with the most repeated ones, since there is only one portfolio. This indicates that stationary relationships between assets appear and disappear, which makes it difficult to manage portfolios based on this methodology.

Based on the above results, we analyze the frequency of participation of each asset in a cointegrating relationship. Table-5 shows the results, excluding those assets that have not participated in any cointegrating relationship over the sample period, and which are: *Total Energies*, *Santander Bank*, *Deustche Telekom*, *AXA*, *Nordea Bank*, and *ENI*.

Table-5 shows that there are assets that are more likely to cointegrate with other assets (regardless of the number of assets that form the cointegrating relationship). In short, cointegrating relationships can be composed of more than two assets; these relationships can be of short or long duration; there are assets more likely to cointegrate with the rest, and the possibility of participating in a cointegration varies according to the number of assets that make up the relationship. All this constitutes an important contribution to the literature, as it allows us to establish asset selection and search criteria before applying the corresponding investment strategy, since, as Bağci and Soylu (2024) showed, a direct relationship between the correlation coefficient and the optimal rebalancing frequency cannot be established.

Table 5. Frequency of participation of each asset in the cointegration relationship.

ASSET	nº Relations	%/total	Frequency
BBVA	12	0.006%	0.006%
SANOFI	16	0.008%	0.014%
ING	47	0.024%	0.038%
MERCEDES	66	0.034%	0.072%
SIEMENS	79	0.040%	0.113%
INTENSA	103	0.053%	0.165%
NOKIA	143	0.073%	0.238%
ALLIANZ	201	0.103%	0.341%
INFINEON	203	0.104%	0.445%
VOLKSWAGEN	351	0.180%	0.625%
DEUSCPOST	1062	0.543%	1.168%
SCHENEIDER	1753	0.897%	2.065%
BMW	1942	0.993%	3.058%
AHOLD	2278	1.165%	4.224%
LVHM	2907	1.487%	5.711%
HERMES	3083	1.577%	7.288%
DANONE	3146	1.609%	8.897%
BASF	3333	1.705%	10.602%
RICARD	4393	2.247%	12.850%
SAP	5049	2.583%	15.433%
STELLANTIS	5068	2.593%	18.025%
DEUSCBORSE	5476	2.801%	20.826%
MUENCHENER	5500	2.814%	23.640%
ESSILOR	5756	2.945%	26.585%
ENEL	6120	3.131%	29.716%
IBERDROLA	6146	3.144%	32.860%
AIRBUS	7105	3.635%	36.494%
LOREAL	8153	4.171%	40.665%
ADIDAS	8160	4.174%	44.840%
KERING	8974	4.591%	49.430%
SAFRAN	9380	4.799%	54.229%
UNICREDIT	9429	4.824%	59.052%
VINCI	10741	5.495%	64.547%
INDITEX	10931	5.592%	70.139%
AIRLIQ	11023	5.639%	75.778%
ANHEUSER	11163	5.711%	81.489%
BAYER	13568	6.941%	88.430%
ASML	22617	11.570%	100.000%
total holdings	195477	100%	

Note: The *total holdings* relationship (195477) is different from *total relationships* (31955) since any asset can be relationships with more than two assets.

Table-6 shows the average results for the cointegrated portfolios grouped by the number of assets involved, including the β of the monthly returns of the cointegrated portfolios with respect to the same return of the market portfolio, taken as the Eurostoxx-50. The portfolios closest to meeting the conditions of statistical arbitrage portfolios are also shown.

Table 6. Average results of cointegrated portfolios by number of assets.

		Average results for all portfolios		
Assets	mean	std dev	coef. Variation	β
2	212.698	162.368	1.310	-1.841
3	438.442	390.838	1.122	1.512
4	451.476	323.019	1.398	-1.086
5	753.858	554.124	1.360	-1.962
6	1247.992	871.209	1.432	1.566
7	1497.635	930.825	1.609	1.018
8	1873.346	1346.020	1.392	1.720
9	2757.653	3832.739	0.719	1.304
10	2645.869	1571.796	1.683	1.017
11	3024.235	2647.695	1.142	2.042
12	2234.612	1235.445	1.809	0.659
13	1042.349	1917.101	0.544	1.055
14	594.917	464.991	1.279	0.335
		Average results for the possible statistical arbitrage portfolios		
Assets	mean	std dev	cv	beta
2	-1.122	2.217	-0.506	-0.381
3	-0.545	4.854	-0.112	-0.393
4	-0.298	4.185	-0.071	0.829
5	-0.049	5.396	-0.009	0.076*
6	-0.647	8.451	-0.077	-0.219
7	-0.056	2.899	-0.019	0.081 *
8	0.584	4.652	0.125	0.175
9	0.125	9.373	0.013	-0.653
10	0.592	4.072	0.145	0.521
11	-1.913	5.952	-0.321	1.290
12	0.116	6.438	0.018	-1.043
13	-2.771	4.352	-0.637	1.457
14	5.917	6.991	0.846	0.335

The results in Table-6 indicate that the mean value of the portfolios is not only conditioned by the number of assets but also depends on the weights (cointegration relationship) and the price of each asset. Note that the standard deviation is high and, in some cases, even higher than the mean value (coefficient of variation less than one). With respect to the portfolios closest to meeting the statistical arbitrage conditions, i.e., self-financed (mean value close to zero) and without the presence of market risk (beta with respect to the Eurostoxx-50 close to zero), we have only been able to find two portfolios (see Table-6 marked with *) for five and seven assets. In both cases, the 5 and 7-asset portfolios are composed of *Asml-Airliq-Airbus-Safran-DeustchBoerse* and *Asml-Vinci-Bayer-Anheuser-Kering-Adidas-Unicredit*, respectively.

Table-7 shows the annualized monthly returns, estimated as a monthly relative rate of change, for the cointegrated portfolios according to the number of assets, the benchmark index (Eurostoxx-50) and the only two portfolios that met the statistical arbitrage requirements. These returns are adjusted with 10 bps as a round-trip transaction cost, as proposed by Tokat and Hayrullahoğlu (2022), although Korniejczuk and Ślepaczuk (2024) the transaction costs have neutral effect (see page 23).

The results in Table-7 show that the cointegration strategies are more profitable than the benchmark, but there is a major drawback when comparing management results from monthly returns, since it is implicitly assumed that each month the portfolios are sold and repurchased, and that is difficult in practice, since, as we found above, cointegration relationships are not stable over time, i.e., they can be short or long term. However, this strategy returns higher profitability than the index market (Eurostoxx-50), even if the portfolios cointegrated in one month are not so in the following month (with the consequent monthly closing position). Finally, while the positive results are in a similar range to those of other studies (Tokat and Hayrullahoğlu (2022), 9%-41%), we found that, unlike these, it is also possible to have negative results.

Table 7. Annualized monthly returns of portfolios with cointegrated assets and adjusted transaction costs.

Assets	Mean	Std. Dev.	Max	Min
2	14.190%	2.7470	38.6%	-28.3%
3	5.968%	2.3365	19.3%	-27.1%
4	-1.889%	2.4914	48.1%	-51.9%
5	1.510%	1.8992	67.3%	-47.8%
6	3.959%	1.4961	44.2%	-23.4%
7	-2.310%	2.4021	68.5%	-46.8%
8	-0.072%	2.2965	23.2%	-30.7%
9	-9.635%	2.9811	46.1%	-55.8%
10	-1.429%	2.5377	28.1%	-53.7%
11	2.932%	1.7520	6.9%	-10.0%
12	7.272%	1.5227	8.8%	-2.9%
13	4.108%	0.7424	35.8%	-5.0%
14	0.858%	0.0563	0.9%	0.9%
EUROSTOXX-50	0.548%	0.0486	18.06%	-16.30%
Arb. Estrat. (5 assets)	2.564%	0.2911	27.36%	-18.92%
Arb. Estrat. (7 assets)	-19.526%	1.5056	33.57%	-21.20%

Additionally, pair trading strategies are based on the stationary behavior of reversion to the mean of the value of the portfolio composed of cointegrated assets, which is identified econometrically with an error correlation model. Under this premise, the logical practice is to buy the portfolio (or take a long position) when it is below its mean value at a certain level (or percentile) and sell it (or take a short position) when the value is above. Thus, Table-8 shows the results of buying and selling the portfolios when they reach a minimum and maximum percentile, respectively. In addition, the total number of cointegration relationships is shown according to the number of assets ($Total\ relat.$). We also show the cointegration relationships by the number of assets for which the hypothesis of a null cointegrating vector is rejected. (*Coint. Relat.* $\beta \neq 0$). The relationships that met (*Buy-Sell*), at some point in the study period, the requirement to buy and sell upon reaching the set percentiles (70% and 90%), the average of the times that buy-sell trade (*trade*) happened, i.e., the lower and upper level was reached (30%-70% or 10%-90%), the average time elapsed to complete the operation (*months*) or months that passed after the value of the portfolio went from one extreme percentile to another with the opposite sign, and finally the average annualized return of trade are also shown.

Table 8. Annualized returns of portfolios with buy-sell strategy and adjusted with transaction cost.

buying and selling decision 30-70% percentile								
Assets	Total relat.	Coint. Relat. $\beta \neq 0$	Buy−Sell	trade	months	mean return	max return	min return
2	2207	2174	528	2	26	27.49%	45.76%	6.33%
3	2557	2339	538	1	30	18.55%	60.30%	3.18%
4	3439	3159	726	1	34	21.17%	31.76%	2.27%
5	4732	4426	950	1	32	25.16%	57.88%	2.23%
6	5853	5524	1325	1	35	20.53%	53.99%	2.57%
7	5929	5536	1337	1	34	19.69%	56.90%	1.94%
8	4642	4331	1084	1	35	23.02%	39.59%	4.59%
9	2808	2618	698	1	36	26.35%	41.16%	3.11%
10	1356	1249	320	1	37	17.57%	61.32%	4.69%
11	513	458	142	1	37	22.19%	59.67%	4.52%
12	161	122	31	1	39	19.45%	31.50%	9.26%
13	24	18	5	1	47	18.54%	45.80%	10.84%
14	2	1	0	0	0	0.00%	0.00%	0.00%
Eurostoxx-50				1	8	16.24%		
arb. Stat. (5)				1	4	21.03%		
arb. Stat. (7)				2	16	20.91%		
buying and selling decision 10-90% percentile								
Assets	Total relat.	Coint. Relat. $\beta \neq 0$	Buy-Sell	trade	months	mean return	max return	min return
2	2207	2174	349	1	40	44.85%	74.90%	17.29%
3	2557	2339	303	1	46	36.15%	68.34%	15.05%
4	3439	3159	417	1	44	35.52%	41.71%	11.47%
5	4732	4426	545	1	43	46.20%	61.57%	11.61%
6	5853	5524	745	1	44	39.28%	58.25%	10.62%
7	5929	5536	801	1	45	37.11%	62.30%	6.70%
8	4642	4331	642	1	46	41.72%	43.77%	18.30%
9	2808	2618	421	1	46	38.51%	52.39%	16.09%
10	1356	1249	172	1	48	32.62%	77.70%	12.06%
11	513	458	85	1	47	29.35%	70.66%	14.31%
12	161	122	22	1	49	30.42%	44.16%	26.51%
13	24	18	4	1	33	28.58%	55.30%	19.98%
14	2	1	0	0	0	0.00%	0.00%	0.00%
Eurostoxx-506			0	0	0.00%			
arb. Stat. (5)			0	0	0.00%			
arb. Stat. (7)6			1	29	32.18%			

From the results in Table-8, we find that not all cointegrating relationships have non-zero cointegrating vectors, while in pairwise relationships 98% of them do; that as the number of assets; in the cointegrated portfolios increases, the percentage is around 90% up to 11 assets and that, from 12 to 14 assets, the percentage is 75%–50%. Note that cointegrated portfolios formed by two assets are not the most common (2,174 cases), while there were 5,536 cases with seven assets. Proportionally (transactions performed versus total cointegrating relationships), for the 70% strategy, only 25% of the cointegration relationships were buy-sell transactions, while for the 90% strategy, that dropped to 15%. This indicates an increase in the difficulty of applying this strategy, since it would not be enough to identify them and check their stability over time; moreover, only a small percentage of the total is profitable. On the other hand, even if there were few profitable relationships, their performance is far superior to that of the benchmark index (Eurostoxx-50), with a lower number of transactions required to achieve it (one buy-sell transaction on average), but with an average waiting time of 36 months (70% strategy) or 48 months (90% strategy), over an analysis period of 121 months. Finally, note that only the statistical arbitrage cointegration relationships are more profitable than the benchmark, with the 7-asset composite standing out, as the trade-off holds for both percentile combinations (70th and 90th). These results reaffirm findings of Bağci and Soylu (2024) on the best long-term statistical arbitrage investment strategy, i.e., buy-to-hold.

In summary, pair trading cannot ignore the fact that cointegrated portfolios can be composed of more than two assets, that these relationships are unstable over time, and that a priori an entry and exit rule must be set. However, once the relationship has been identified and the criterion (percentile) has been met, one would only need to wait for the reversal process in order to achieve significant returns with few operations, which reduces transaction costs.

3.2. Robustness analysis

In this section, we complete our study with a series of robustness analyses. First, to test the statistical consistency of the sample size in each of the monthly estimates, we run a regression as follows:

$$nr_i = \alpha_0 + \alpha_1 \cdot T_i + \alpha_2 \cdot T_i^2 + u_i \tag{3}$$

where T_i is the sample size used to perform the estimate i, while nr_i is the number of cointegration relationships resulting in i-estimate or month. Recall that each estimate has one more observation than the previous one, i.e., the first estimate is done with 156 monthly observations and the last one with 276, so that 121 monthly estimates are done. Note that if the parameters associated with the sample size are not statistically significant, then the estimates obtained are independent of the sample size.

Table 9. Relationship between sample size and results.

Parameters	Value	p-value
constant	-7263.4265	0.0791
T	69.4270	0.0951
T^2	-0.1560	0.1831
R^2	2.95%	

The results in Table-9 clearly show that the sample size is not significant in explaining the number of cointegrating relationships found and justifies the findings of Tokat and Hayrullahoğlu (2022) on high sensitivity of the results to the window (70%-93% in Table-3).

In addition, we analyze the robustness of our proposal to determine the set of possible cointegration relationships:

1. We estimate the eigenvalues of the correlation matrix between the cointegrating relationships found each month (since the relationships change in each estimate month) as a function of the

number of assets involved in the cointegration relationship to determine the relationship between the cointegrated portfolios. Thus, Table-10 shows the number of eigenvalues needed to explain the correlation matrix between cointegrated portfolios for each month to at least 95% and how many times this number of eigenvalues is repeated. In light of the results of Table-10, we can see that with a maximum of six factors, this matrix is explained in all cases, irrespective of the number of assets participating in the relationship[6]. For most cases, between two and five factors are required. This is relevant because it indicates that most cointegration relationships are linear combinations of others, since with a small number of factors we can explain most of their correlation matrices. This, therefore, endorses the initial assumption of the methodological proposal, since from lower-order relationships (starting from pairs), we analyze higher order portfolios. In addition, although not included in our objective, it opens up the opportunity to analyze whether these factors correspond to those usually used in asset pricing models.

Table 10. Principal Components Analysis on correlation matrix from cointegrated portfolios.

PC	Asset-2	Asset-3	Asset-4	Asset-5	Asset-6	Asset-7	Asset-8	Asset-9	Asset-10	Asset-11	Asset-12	Asset-13
1	2	0	0	0	0	0	0	0	0	0	0	0
2	8	35	27	25	22	11	9	5	4	3	0	0
3	25	38	59	43	26	19	14	9	5	3	1	1
4	50	34	19	22	18	17	11	7	4	3	2	1
5	36	12	2	6	11	9	6	4	3	1	0	0
6	0	0	0	1	2	3	0	0	0	0	0	0
7 o more	0	0	0	0	0	0	0	0	0	0	0	0
total	121	119	107	97	79	59	40	25	16	10	3	2

Note: *PC* is the minimum number of factors or eigenvalues to achieve a capacity explanatory of the correlation matrix of at least 95%; *total* is the number of months in which cointegration relationships have been found with the respective number of assets (see Table-3).

Table 11. Relationship of co-integrated portfolios with more than 2 assets and co-integrated portfolios in pairs.

Assets	*min relations*	*Max relations*	*min* $Adj.R^2$	*Max* $Adj.R^2$
3	8	162	80.81%	99.05%
4	6	482	81.65%	98.63%
5	2	1102	84.03%	98.91%
6	3	1886	83.97%	96.40%
7	8	2392	81.53%	97.47%
8	3	2283	82.62%	95.82%
9	2	1598	83.55%	95.95%
10	2	894	79.99%	96.59%
11	1	354	80.80%	97.19%
12	1	94	81.74%	96.51%
13	1	17	80.15%	93.54%
14	1	1	78.37%	78.37%
45 (all)	22	43	77.06%	97.63%

[6]Except in the case of 14 assets, which is only found on one occasion, it is not possible to estimate the correlation matrix to perform a principal component analysis.

2. We run the following regression:

$$x_{k,i,t} = \omega_0 + \sum_{j=1}^{J} \omega_{j,1} \cdot x_{2,j,t} + u_{k,i,t} \tag{4}$$

where, for each monthly estimation t, $x_{k,i,t}$ is the standardized value of cointegration portfolio i composed of $k > 2$ assets, while $x_{2,j,t}$ is the standardized value of cointegration portfolio j composed of two assets. Table-11 contains the minimum and maximum number of cointegration relationships found for each grouping of assets higher than 2 and even for the total assets as a whole, and also shows the maximum and minimum values of the adjusted R^2 resulting from expression-4.

Based on Table-11 results, the explanatory power of cointegrated asset pairs is very high, so that our assumption about their linear combination to form cointegration relationships with a larger number of assets cannot be rejected. Recall the case of the portfolio of 14 assets that only appears once in the analysis period.

3. We analyze a possible cause of the instability of the cointegration relationships found. To do so, we run the following regression:

$$\begin{aligned} z_t &= \lambda_0 + \lambda_1 \cdot p_t + \lambda_2 \cdot p_t^2 + e_t \\ z_t &= \lambda_0 + \lambda_1 \cdot p_t + \lambda_2 \cdot p_t^2 + \lambda_3 \cdot D_t + v_t \end{aligned} \tag{5}$$

where z_t is the number of cointegration relationships composed of more than two assets, p_t is the number of cointegrating relationships with only two assets, and D_t is a dummy variable which takes the value 1 if the asset participating in the highest number of pairwise relationships is the same as in the previous period, and takes the value 0 otherwise. Table-12 shows the results.

Table 12. Relationship between the number of co-integrated portfolios with 2 assets and the number of co-integrated portfolios with more than 2 assets.

Parameters	Value	p-value
	Panel A. Without dummy	
constant	43.3791	0.004
number pair trading	-87.7725	0.000
number pair trading^2	3.2748	0.000
R^2	50.28%	
	Panel B. With dummy	
constant	35.8343	0.068
number pair trading	-88.6394	0.000
number pair trading^2	3.3012	0.000
dummy	-29.8539	0.031
R^2	55.94%	

From Table-12, note that the number of pairwise relationships explains in a non-linear way the number of higher order relationships found, so we again validate the assumption of our bottom-up proposal. We also observe that when there is a change in the asset with the highest pairwise cointegration, there is a linear decrease in the higher order relationships (approximately 30 fewer relationships), although the total number of relationships also depends on the pairwise relationships in which this asset participates and their square.

4. Conclusions and Discussion

Statistical arbitrage aims to take advantage of the reversion to mean value of asset portfolios by following their long-term trend and, within the approaches using cointegration techniques, the Johansen cointegration analysis procedure is the most consistent.

Unlike previous studies, in this empirical analysis, we go one step further, since we do not settle for asset-pair strategies, but look for any combination of assets that presents only one cointegration relationship, regardless of the number of assets involved. To do so, we developed an algorithm to search for those relationships and, starting from the lower order relationships (pairs) and increasing the number of assets that can form a cointegrated portfolio between assets previously cointegrated in the previous order, we look not only for pair trading but also for linear combinations of the set of lower order relationships. We expand the set of possible investments to find the most profitable one and closest to the requirements of statistical arbitrage (self-funded and zero market beta).

We applied our methodological proposal to a set of monthly prices of 45 assets included in the Eurostoxx-50 from January 2002 to December 2024. and the results obtained show that the cointegration relationships are not stable over time, i.e., there are moments in which there are numerous relationships (more than 3,000) compared to other months in which there are barely four relationships. Moreover, these relationships were observed to be very volatile. We also found that there are some assets that never participate in cointegration relationships. Regarding the number of assets involved in each cointegration relationship, we found empirical evidence that there are portfolios of two (pair trading) to 14 assets. Moreover, pair trading is not the most frequent. As for the portfolios that meet the statistical arbitrage portfolio criteria, we only found two possible portfolios composed of five and seven assets.

The performance of these strategies in terms of monthly return (assuming that the portfolio is only held for one month) is superior to that of the Eurostoxx-50 benchmark for all possible combinations (from two to 14 assets). When we analyze the strategy in terms of reversion to the mean with an invest and disinvest criterion based on the historical percentiles of the values of each cointegration portfolio (70th and 90th), we observe that the cointegrated portfolios move from one end of the distribution to the other very infrequently, for which an average period of around 3.5 years must elapse, although the capital gain and the return obtained are high. It would therefore involve a strategy of identifying stable cointegration relationships, waiting to take the position when its value touches the distribution tail and waiting again until it reaches the opposite tail, receiving high remuneration (long-term passive strategy) without high transaction costs. However, a shorter term strategy for taking a position in the cointegrated portfolios for one month and closing the position the next month could also be carried out. This has also been shown to be more profitable than the market index but would entail higher transaction costs.

We also found that at each moment in time, the assets with the highest participation in cointegration relationships are different, and we have shown the instability of cointegration relationships over time. Through robustness analysis, we found that the number of cointegrating relationships depends on the pairwise relationships found and also on the change of asset that participates the most in these pairwise relationships (pivot asset).

The results obtained are important for future research, and even open the door to future work analyzing the explanatory factors of the volume of cointegrated portfolios. They will also be useful for the financial industry and investors in general since they provide insights into the overall behavior of the groups of assets traded on the market and help to find and monitor the assets that are most involved in cointegration relationships and thus in long-term trends.

Use of AI tools declaration

The authors declare they have not used Artificial Intelligence (AI) tools in the creation of this article.

Author contributions

González-Sánchez, Mariano: Conceptualization, Methodology, Software, Data curation, Writing-Original draft preparation, Visualization, Investigation, Supervision, Software, Validation, Writing-Reviewing and Editing.

Nave Pineda, Juan M.: Conceptualization, Methodology, Software, Data curation, Writing- Original draft preparation, Visualization, Investigation, Supervision, Software, Validation, Writing- Reviewing and Editing.

Ruiz Rincón, Javier: Conceptualization, Methodology, Software, Data curation, Writing- Original draft preparation, Visualization, Investigation, Supervision, Software, Validation, Writing- Reviewing and Editing,

Acknowledgments

This work is finnancied by: Grant PID2023-151010NB-I00 funded by MICIU/AEI/10.13039/501100011033 and FEDER-EU.

Conflict of interest

All authors declare no conflicts of interest in this paper.

References

Bağci M, Soylu PK (2024) Classification of the optimal rebalancing frequency for pairs trading using machine learning techniques. *Borsa Istanb Rev* 24: 83–90. https://doi.org/10.1016/j.bir.2024.12.004

Bilgili F (1998) Stationarity and cointegration tests: Comparison of Engle-Granger and Johansen methodologies. MPRA Paper 75967, University Library of Munich. Available from: https://mpra.ub.uni-muenchen.de/75967/.

Bondarenko O (2003) Statistical Arbitrage and Securities Prices. *Rev Financ Stud* 16: 875–919. https://doi.org/10.1093/rfs/hhg016

Bowen D, Hutchinson MC, O'Sullivan N (2010) Frequency Equity Pairs Trading: Transaction Costs, Speed of Execution, and Patterns in Returns. *J Trading* 5: 31–38. https://doi.org/10.3905/jot.2010.5.3.031

Cavaliere G, Rahbek A, Taylor AMR (2010) Cointegrarion rank testing under conditional heteroskedasticity. *Econom Theory* 26: 1719–1760. https://doi.org/10.1017/S0266466609990776

Chan F (2013) Stationarity and cointegration tests: Comparison of Engle-Granger and Johansen methodologies. SSRN Working Paper number 2223753. https://doi.org/10.2139/ssrn.2223753

Chen HJ, Chen SJ, Chen Z, et al. (2019) Empirical Investigation of an Equity Pairs Trading Strategy. *Manage Sci* 65: 370–389. https://doi.org/10.1287/mnsc.2017.2825

Doornik JA (1998) Approximations to the Asymptotic Distribution of Cointegration Tests. *J Econ Surv* 12:573–593. https://doi.org/10.1111/1467-6419.00068

Elliott RJ, Van Der Hoek J, Malcolm WP(2005) Pairs trading. *Quant Financ* 5: 271–276. https://doi.org/10.1080/14697680500149370

Engle RF, Granger CWJ (1987) Co-Integration and Error Correction: Representation, Estimation, and Testing. *Econometrica* 55: 251–275. https://doi.org/10.2307/1913236

Ferré M, Hall SG (2002) Foreign exchange market efficiency and cointegration. *Appl Financ Econ* 12: 131–139. https://doi.org/10.1080/09603100110090055

Ferré M (2004) The Johansen Test and the Transitivity Property. *Econ Bull* 3: 1–7.

Gatev E, Goetzmann WN, Rouwenhorst KG (2006) Pairs Trading: Performance of a Relative-Value Arbitrage Rule. *Rev Financ Stud* 19: 797–827. https://doi.org/10.1093/rfs/hhj020

Haddad GK, Talebi H (2023) The profitability of pair trading strategy in stock markets: Evidence from Toronto stock exchange. *Int J Financ Econ* 28: 193–207. https://doi.org/10.1002/ijfe.2415

Hanson TA, Hall J (2012) Statistical Arbitrage Trading Strategies and High Frequency Trading. Working Paper number 2147012, SSRN. https://doi.org/10.2139/ssrn.2147012

Hogan S, Jarrow R, Teo M, et al. (2004) Testing market efficiency using statistical arbitrage with applications to momentum and value strategies. *J Financ Econ* 73: 525–565. https://doi.org/10.1016/j.jfineco.2003.10.004

Johansen S (1991) Estimation and Hypothesis Testing of Cointegration Vectors in Gaussian Vector Autoregressive Models. *Econometrica* 59: 1551–1580. https://doi.org/10.2307/2938278

Kim K (2003) Dollar exchange rate and stock price: evidence from multivariate cointegration and error correction model. *Rev Financ Econ* 12: 301–313. https://doi.org/10.1016/S1058-3300(03)00026-0

Krauss C, Stübinger J (2017) Non-linear dependence modelling with bivariate copulas: statistical arbitrage pairs trading on the S&P 100. *Appl Econ* 49: 1–18. https://doi.org/10.1080/00036846.2017.1305097

Krauss K (2017) Statistical arbitrage pairs trading strategies: Review and outlook. *J Econ Surv* 31: 513–545. https://doi.org/10.1111/joes.12153

Korniejczuk A, Ślepaczuk R (2024) Statistical arbitrage in multi-pair trading strategy based on graph clustering algorithms in US equities market. https://arxiv.org/abs/2406.10695

Kurita T (2013) Exploring the Impact of Multivariate GARCH Innovations on Hypothesis Testing for Cointegrating Vectors. *Commun Stat Simul C* 42: 1785–1800. https://doi.org/10.1080/03610918.2012.677920

Lee H, Lee J, Im K (2015) More powerful cointegration tests with non-normal errors. *Stud Nonlinear Dyn Econom* 19: 397–413. https://doi.org/10.1515/snde-2013-0060

Lee H, Hur M (2021) Non-normal errors or nonlinearity? Performance of unit root tests. *Appl Econ* 53: 6094–6103. https://doi.org/10.1080/00036846.2021.1937038

Östermark R (2001) Multivariate cointegration analysis of the Finnish-Japanese stock markets. *Eur J Oper Res* 134: 498–507. https://doi.org/10.1016/S0377-2217(00)00272-1

Perlin M (2007) M of a Kind: A Multivariate Approach at Pairs Trading. Working paper 952782, SSRN. https://doi.org/10.2139/ssrn.952782

Rad H, Low RKY, Faff R (2016) The profitability of pairs trading strategies: distance, cointegration and copula methods. *Quant Financ* 16: 1541–1558. https://doi.org/10.1080/14697688.2016.1164337

Ramos-Requena JP, Trinidad-Segovia JE, Sánchez-Granero MA (2017) Introducing Hurst exponent in pair trading *Physica A* 488: 39–45. https://doi.org/10.1016/j.physa.2017.06.032

Ramos Jungblut J (2024) Multivariate Cointegration in Statistical Arbitrage. Working Paper number 4906546, SSRN. https://doi.org/10.2139/ssrn.4906546

Sheng HC, Tu AH (2000) A study of cointegration and variance decomposition among national equity indices before and during the period of the Asian financial crisis. *J Multinatl Financ Manag* 10: 345–365. https://doi.org/10.1016/S1042-444X(00)00034-7

Silvapulle PS, Podivinsky JM (2000) The effect of non-normal disturbances and conditional heteroskedasticity on multiple cointegration tests. *J Stat Comput Simul* 65: 173–189. https://doi.org/10.1080/00949650008811997

Stander Y, Marais D, Botha I (2013) Trading strategies with copulas. *J Econ Financ Sci* 6: 83–108. https://doi.org/10.4102/jef.v6i1.278

Stock JH (1987) Asymptotic Properties of Least Squares Estimators of Cointegrating Vectors. *Econometrica* 55: 1035–1056. https://doi.org/10.2307/1911260

Stock JH, Watson MW (1993) Simple estimator of cointegrating vectors in higher order integrated systems. *Econometrica* 61: 783–820. https://doi.org/10.2307/2951763

Tokat E, Hayrullahoğlu AC (2022) Pairs trading: is it applicable to exchange-traded funds? *Borsa Istanb Rev* 22: 743–751. https://doi.org/10.1016/j.bir.2021.08.001

Vidyamurthy G (2004) *Pairs Trading, Quantitative Methods and Analysis*. John Wiley & Sons, Hoboken, NJ.