Metadata of the article that will be visualized in OnlineFirst

ArticleTitle	Pre-selection in cointegration-based pairs trading
Article Sub-Title	
Article CopyRight	The Author(s) (This will be the copyright line in the final PDF)
Journal Name	Statistical Methods & Applications
Corresponding Author	FamilyName **Brunetti** Particle Given Name **Marianna** Suffix Division Department of Economics and Finance Organization University of Rome Tor Vergata, CEFIN & CEIS Address Office 27, First Floor, Research Building, Rome, Italy Phone Fax Email marianna.brunetti@uniroma2.it URL ORCID http://orcid.org/0000-0002-5332-1348
Author	FamilyName **Luca** Particle De Given Name **Roberta** Suffix Division Organization Bank of Italy Address Rome, Italy Phone Fax Email URL ORCID
Schedule	Received Revised Accepted 10 Apr 2023
Abstract	The paper compares the final profitability of a cointegration-based pairs trading strategy when pairs of stocks are pre-selected by means of seven different measures. Some of the measures considered have been extensively used in the pairs trading literature, while others represent a novelty in this type of application. We find that pre-selection matters, since the excess returns remarkably vary, in terms of both average and variability, depending on the metrics used. Differences in profitability by pre-selection metrics are retrieved even after considering commissions and cut rules, market impact, a stricter definition of the Spread reversion to the equilibrium and alternative cointegration tests. Besides, the pairs trading profitability is found to be heterogeneous across the different pre-selection metrics also in terms of exposure to the systematic stock-market risk factors.
Keywords (separated by '-')	Pairs trading - Pre-selection - Cointegration - Spectral coherence - Risk factors
JEL Classification	G10 - C40 - C50
Footnote Information	

Journal : SmallExtended 10260	Article No : 702	Pages : 32	MS Code : 702	Dispatch : 21-4-2023

Statistical Methods & Applications
https://doi.org/10.1007/s10260-023-00702-4

ORIGINAL PAPER

Pre-selection in cointegration-based pairs trading

Marianna Brunetti[1] · **Roberta De Luca**[2]

Accepted: 10 April 2023

Abstract

The paper compares the final profitability of a cointegration-based pairs trading strategy when pairs of stocks are pre-selected by means of seven different measures. Some of the measures considered have been extensively used in the pairs trading literature, while others represent a novelty in this type of application. We find that pre-selection matters, since the excess returns remarkably vary, in terms of both average and variability, depending on the metrics used. Differences in profitability by pre-selection metrics are retrieved even after considering commissions and cut rules, market impact, a stricter definition of the Spread reversion to the equilibrium and alternative cointegration tests. Besides, the pairs trading profitability is found to be heterogeneous across the different pre-selection metrics also in terms of exposure to the systematic stock-market risk factors.

Keywords Pairs trading · Pre-selection · Cointegration · Spectral coherence · Risk factors

JEL Classification G10 · C40 · C50

1 Introduction

Firstly introduced in the '80s by Gerry Bamberger and the quantitative group led by Nunzio Tartaglia's at Morgan Stanley (Bookstaber 2007), the pairs trading strategy has since then became very popular. The idea behind pairs trading strategy is rather simple and consists of identifying pairs of assets whose prices share a common stochastic trend, and of profiting by exploiting any deviation from this long-term relationship, which, by construction, is temporary. More specifically, whenever the prices of these assets diverge from their long-run equilibrium,

✉ Marianna Brunetti
marianna.brunetti@uniroma2.it

1 Department of Economics and Finance, University of Rome Tor Vergata, CEFIN & CEIS, Office 27, First Floor, Research Building, Rome, Italy

2 Bank of Italy, Rome, Italy

the relatively overpriced asset is sold and the other asset is bought so that when prices converge again, by unwinding the positions, the profit is realized. Pairs trading is a statistical arbitrage strategy (Bondarenko 2003) because, although a positive payoff may not be guaranteed in each state, its expected value is positive. Moreover, it is expected to be a market-neutral strategy (Ehrman 2006) because its results depend on the relative pricing of the selected assets only, so that the potential profits are independent from market performance.

The first step to take for pairs trading implementation is to identify pairs of assets. Among the different methodologies developed, the one based on cointegration tests has certainly attracted great attention due to its superior results in terms of profitability (Huck and Afawubo 2015; Rad et al. 2016; Blázquez et al. 2018). However, as confirmed e.g. in Huck and Afawubo (2015), its application to large datasets (and ideally to the entire markets) comes with a remarkable computational burden. For example, a dataset of 500 assets would require 124.750 cointegration tests to identify all pairs that are potentially eligible for trading. To overcome this drawback, some empirical contributions narrow down the analysis to a subset of assets in the market, previously selected based on some measures of distance or association. If, on the one hand, this allows to reduce considerably the computational burden of carrying out the first step of cointegration-based pairs trading, on the other hand, the effect of this pre-selection of pairs on the final profitability of the strategy is not clear and has so far never been assessed. The first contribution of this paper is thus to fill this gap, by investigating and comparing the profitability of a cointegration-based pairs trading strategy when pairs are pre-selected based on seven different pre-selection measures.

Some of the measures considered for pre-selection have been extensively used in the pairs trading literature, either to pre-select or to identify assets' couples (rather than via cointegration tests). This is the case for the correlation between the log-prices, the correlation between the returns, the sum of squared deviations, and the price-ratio between the normalized log-prices. These measures are of easy interpretation and fast computation, and hence particularly appreciated by practitioners for whom speed and efficiency of computation is a vital consideration (Clark 2012; Brogaard et al. 2014; Angel 2014). However, they all present some drawbacks. These include: (i) the potential risk of pre-selecting pairs of assets whose prices (or returns) display low volatilities, and (ii) they are not able to detect the common trend between the paired assets. The set of metrics considered for pre-selection is thus augmented with the covariance between the log-prices and the covariance between the returns, which overcome the first weakness, and with the spectral coherence at frequency zero, which specifically addresses the latter issue. These additional measures used for pre-selection represent a novelty in this type of application and thus constitute the second contribution of this paper.

The rest of paper is structured as follows: Sect. 2 briefly summarizes the empirical contributions dealing with distance-based and cointegration-based pairs trading; Sect. 3 presents the dataset and the methodology used, illustrating the pre-selection measures considered, the trading strategy, and its final evaluation; Sect. 4 presents the main results, while Sect. 5 extends the analysis by investigating the risk-profiles of the excess returns and tests the robustness of the main results. Finally, last Section concludes.

2 Literature

Pairs trading strategies have been implemented using different approaches, which Krauss (2017) broadly classifies into the following categories: (i) distance approach; (ii) cointegration approach; (iii) time series (or stochastic spread) approach; and (iv) a residual category, gathering all the applications not belonging to one of the above. All the approaches require to first identify pairs of assets (during the so-called formation period) and then to implement an investment strategy (during the so-called trading period). The different approaches mainly differ for how pairs are selected during the formation period. For instance, in the distance approach pairs of assets are identified using nonparametric distance metrics, while in the cointegration approach Engle and Granger (1987) and/or Johansen (1988) cointegration tests are employed to spot a long-term equilibrium between asset prices time series.[1]

Gatev et al. (2006) are among the first applying pairs trading using the distance approach. As the authors outline, the implementation of such strategy is structured in two periods: (1) in the formation period, pairs are selected by minimizing the sum of squared deviations between their normalized prices[2]; and (2) in the trading period, a position is opened whenever the distance between the normalized prices diverges more than a given threshold. The investment strategy consists in selling one dollar of the relatively overpriced asset and buying one dollar of the underpriced asset, thereby ensuring that the strategy is self-financing. The position is then closed if and when the normalized prices difference reaches zero, or at the end of the trading period. Since the strategy is self-financing, all the final payoffs can be interpreted as excess returns.

In an analysis on the liquid US stocks in the CRSP, over the period 1962 to 2002, Gatev et al. (2006) investigate the profitability of the distance-based pairs trading strategy and prove that it produces significant excess returns, which also survive to the inclusion of trading costs. The very same approach is implemented in many other empirical contributions, including e.g. Do and Faff (2010) and Huck (2013). The former show that the profitability found by Gatev et al. (2006) reduces if the analysis is extended to 2009, while the latter investigates the sensitivity of the profitability found by Gatev et al. (2006) under different parameterizations in terms of the length of the formation period and of the opening threshold. Other contributions rely on pairs trading implemented with the distance approach to investigate the potential sources of profitability of this strategy. Examples include Andrade et al. (2005), who prove the relevance of uninformed demand shocks, or Papadakis and Wysocki (2007) and Engelberg et al. (2009), both focusing on the impact of news disclosure or information events on the strategy performance. Jacobs and Weber (2016) explore the implications of time-varying awareness of firm-level information

1FL01 [1] Being the most relevant for this work, our focus will be on distance and cointegration approaches only.

2FL01 [2] The normalization is performed scaling both log-prices time series to start at 1$.

on asset pricing, proving that pairs opening in high distraction days are more likely to generate higher returns. Regardless of the aim of the study, all these contributions identify pairs by minimizing the sum of squared deviations. The only one who, to the best of our knowledge, uses a different metrics is Chen et al. (2017). They use the Pearson correlation coefficient between returns, in univariate and quasi-multivariate settings, and select, for each asset, the 50 most correlated stocks to form an equally weighted portfolio. As reported by Krauss (2017), their empirical application, conducted on CRSP data between 1962 and 2002, shows that the quasi-multivariate setting is more profitable with respect to the univariate case, and that the profits of the latter are slightly lower if compared to the Gatev et al. (2006) strategy.

As Do et al. (2006) highlight, identifying pairs of assets by minimizing the sum of squared deviations between their prices is proven profitable in different markets, is easy to implement and is model-free. However, as suggested by Krauss (2017), this approach appears to be suboptimal because it can lead to the selection of pairs with a low level of variance in their price distance, thus, de facto, reducing the number of potential trade opportunities. Moreover, this approach is not able to capture the long-run equilibrium relationship between prices and may not guarantee the mean reversion of their differences—a limitation somewhat overcome using the cointegration approach.

In the cointegration-based pairs trading (Vidyamurthy 2004), pairs are identified based on cointegration tests, so as to exploit the mean-reversion property of a stationary process. The underlying idea is that if the prices series of two stocks are cointegrated, there exists a stationary linear combination of the two series that is mean-reverting, meaning that any short-term deviation from this equilibrium is temporary by construction. Similar to the distance approach, the implementation of cointegration-based pairs trading requires two stages: (1) during the formation period, the pairs of assets whose prices are cointegrated are selected; (2) during the trading period, a self-financing strategy is implemented on the identified pairs, buying 1\$ of the relatively low-priced asset and selling 1\$ of the relatively over-priced asset every time the stationary linear combination of the two series, named Spread, sufficiently deviates from its long-run mean.

The first empirical applications of cointegration-based pairs trading focused on commodities futures and spot prices. For instance, Wahab and Cohn (1994) applied it to gold and silver cash and future prices and Girma and Paulson (1999) on the crack spread, which is the prices' difference between petroleum and refined products futures. Similarly, Simon (1999) focused on the crush spread, that is the soybean futures and its end products prices' difference, and Emery and Liu (2002) on the spark spread, i.e. difference between natural gas and electricity futures prices. The same approach has been more recently used e.g. by Gutierrez and Tse (2011), who use CSRP data on three water utility stocks and prove that most of the pairs trading profits are obtained from the Granger-follower position. Applications to the stock market include the contributions by Dunis et al. (2010), who use (daily and intra-daily) data of EuroStoxx 50 index constituents, and by Caldeira and Moura (2013), who apply it to the 50 most liquid stocks of the Brazilian Ibovesoa index. Hence, the application of this approach to larger datasets (and ideally to the entire market) is actually scant. Indeed, the high computational cost of the cointegration-based pairs trading makes its application to large datasets very difficult and explains the typical focus in the empirical literature on small sets of assets.

Some recent contributions have attempted to reduce the computational burden entailed by cointegration tests by pre-selecting assets before testing for cointegration. The only examples in this direction we are aware of are Miao (2014), Huck and Afawubo (2015) and Rad et al. (2016). Miao (2014) proposes to rank pairs of stocks based on the prices' Pearson correlation coefficient, and to test for cointegration only those with correlation at least equal to 0.9. In this way, despite the empirical application uses data on 177 energy companies stocks traded in NTSE and NASDAQ markets, the actual number of cointegration tests required to implement the pairs trading is sensibly reduced, from 15,576 potential pairs to (an average of) 1,378 actually tested pairs. Of those, the first 10 pairs with smallest residuals ADF test statistic are considered eligible for trading, and their final performance is evaluated using the Sharpe ratio. The empirical application in Huck and Afawubo (2015) relies on a sub-sample of the S&P 500 index constituents. Among the 500 stocks, only the pairs of assets whose returns differ no more than 10% are included in the sample and then tested for cointegration. This allows to sensibly reduce the actual number of pairs tested since approximately 80% of the pairs are dropped before testing for cointegration. Finally, Rad et al. (2016) apply a cointegration-based pairs trading to a large dataset composed by 23.616 stocks in CRSP, from 1962 to 2014. In the empirical work, pairs are first sorted based on the sum of squared deviations between prices and, then, cointegration tests are performed until 20 cointegrated pairs are identified.

To be noticed that each of the above cited contribution uses a different pre-selection measure. However, no evidence has so far been provided on the differences, if any, in terms of final profitability and risk-exposure of the cointegration-based pairs trading strategy when assets are pre-selected with different metrics. The first contribution of this paper is thus to fill this gap. Besides, we provide evidence related to three supplementary metrics that, to the best of our knowledge, have never been used in this type of applications, but whose characteristics might help to overcome some of the drawbacks entailed by the measures used so far in the literature. In doing so, we will rely on data referred to the US stock market, as detailed in the following Section.

3 Data and methodology

The empirical analysis relies on the dividend adjusted daily closing prices of the S&P 500 index constituents, which are extremely liquid assets, characterized by high market capitalization, and relatively low transactions costs.The data, retrieved from Thomson Reuters DataStream, cover the period from 1st January 1998 until 30th October 2018, and include all the stocks belonging to the S&P 500 on the last day of our sample (since some stocks are not included in the index all along the sample, the total number of stocks varies between 373 and 505).

Based on this dataset, our empirical application proceeds as follows:

Pairs pre-selection: we consider a one-year formation period during which we order pairs of stocks according to seven different metrics, described in Sect. 3.1;

Cointegration-based pairs identification: using the formation period data and following the ranking as from step 1, we run the cointegration tests required to find the first 20 cointegrated pairs of stocks. This allows us to estimate the parameters and assess the stationarity of the cointegration relationship, both required for the subsequent implementation of the trading strategy. This step is described in greater detail in Sect. 3.2;

Pairs trading: using data from a six-month trading period, we implement the trading strategy described in Sect. 3.3;

Profits evaluation: we compute the monthly excess profits on the six-months trading period and repeat this procedure every month in a rolling window setting. We then assess the profitability of the pairs trading strategy, as described in Sect. 3.4.

3.1 Pre-selection measures

We consider seven different measures to pre-select assets. The first four have been extensively used in the pairs trading literature, either to pre-select or to identify assets to be traded. The last three, which conversely represent a novelty in this type of application, are examined in the light of their features, that are potentially able to overcome some of the drawbacks typical of the first set of measures considered. The importance of considering such a disparate set of metrics to pre-select pairs is confirmed by the results reported in Sect. 3.2 below.

The time-varying volatility characterizing the returns time-series might represent an issue at this stage. In order to overcome this potential drawback, pairs pre-selection is performed on the returns (and the associated log-prices) only after having "cleaned out" their heteroskedasticity, modelled through an exponentially weighted moving average.[3] Specifically, the "homoskedastic" returns, indicated in what follows with ${}_h r_t$, are obtained as:

$$ {}_h r_t = r_t / \hat{\sigma}_t $$

with $t = 1, \ldots, T$, where T is the number of trading days comprising the formation period, and $\hat{\sigma}_t$ is the standard deviation of the raw log-returns r_t. In turn, $\hat{\sigma}_t$ is modelled as:

$$ \hat{\sigma}_t = \sqrt{k\hat{\sigma}^2_{t-1} + (1-k)\hat{u}^2_t} $$

[3] A formal treatment of time-varying volatility, e.g. via a GARCH model, would have considerably increased the computational time required for the pre-selection stage, thus jeopardizing the final aim of the procedure.

where $\hat{u}_t = r_t - \bar{r}$ is the demeaned return at time t, the parameter k is comprised between 0 and 1, and $\hat{\sigma}_0^2$ is a required initial condition.[4] The "homoskedastic" log-prices are derived cumulating the above defined ${}_hr_t$.

The first measure considered for pre-selection is by far the most widely applied in the pairs trading literature. The Sum of Squared Deviations between the normalized "homoskedastic" log-prices was firstly proposed as a selection criterion by Gatev et al. (2006) and then employed for pre-selection of pairs by Rad et al. (2016). It is computed as:

$$\widehat{SSD} = \sum_{t=1}^{T} \left({}_h\tilde{p}_{1,t} - {}_h\tilde{p}_{2,t}\right)^2 \tag{1}$$

where ${}_h\tilde{p}_{1,t}$ and ${}_h\tilde{p}_{2,t}$ are the normalized "homoskedastic" log-prices of stock 1 and 2 on day t, respectively, i.e. ${}_h\tilde{p}_{1,t} = {}_hp_{1,t}/{}_hp_{1,t-1}$ and ${}_h\tilde{p}_{2,t} = {}_hp_{2,t}/{}_hp_{2,t-1}$.

The second measure considered is the Price Ratio, that is:

$$\widehat{PR} = \frac{1}{T}\sum_{t=1}^{T} \frac{{}_h\tilde{p}_{1,t}}{{}_h\tilde{p}_{2,t}} \tag{2}$$

where ${}_h\tilde{p}_{1,t}$ and ${}_h\tilde{p}_{2,t}$ are the above-defined normalized "homoskedastic" log-prices of stock 1 and 2 on day t, respectively. This measure was first proposed for pairs identification by Baronyan et al. (2010), who use the test of price ratio stationarity as an alternative to the test for cointegration.

The third measure considered is the absolute value of the Pearson correlation between the "homoskedastic" log-prices time series, employed in Miao (2014), that is:

$$\hat{\rho}^p = \left| \frac{\sum_{t=1}^{T}\left[{}_hp_{1,t} - {}_h\bar{p}_1\right]\left[{}_hp_{2,t} - {}_h\bar{p}_2\right]}{\sqrt{\sum_{t=1}^{T}\left[{}_hp_{1,t} - {}_h\bar{p}_1\right]^2 \sum_{t=1}^{T}\left[{}_hp_{2,t} - {}_h\bar{p}_2\right]^2}} \right| \tag{3}$$

where ${}_hp_{1,t}$ and ${}_hp_{2,t}$ are the "homoskedastic" log-prices of stock 1 and 2 on day t, ${}_h\bar{p}_1$ and ${}_h\bar{p}_2$ are their corresponding sample means over the formation period. The impact of using this measure to pre-select assets on the final profitability of pairs trading is not clear a priori. On the one hand, it is true that this measure is not directly linked to cointegration, as high correlation might be observed even when cointegration is

4FL01 [4] The parameter k and the initial condition $\hat{\sigma}_0^2$ are defined optimally, i.e. minimizing $\sum_{t=1}^{T}\left(\hat{u}_t^2 - \hat{\sigma}_{t-1}^2\right)^2$.
4FL02 The estimates of $\hat{\sigma}_0^2$ across the 505 assets in the dataset ranges between 0.0028% and 31.07%, with an
4FL03 average of 6.67% and a standard deviation of 5.47% (95% of the estimates are below 19%). As for the
4FL04 smoothing parameter k, the estimates range between 0.0001 and 0.4637, with an average of 0.0477 and a
4FL05 standard deviation of 0.0377 (95% of the estimates are below 0.10). As a robustness check we repeat the
4FL06 exercise imposing k = 0.06, as in Risk Metrics, obtaining very similar results, not reported for reasons of
4FL07 space but available upon request.

absent.[5] On the other hand, Miao (2014) argues that coupling cointegration-based pairs trading with pre-selection based on correlation might be beneficial to trading as they provide different and potentially complementary information. Indeed, while correlation captures co-movements which may be unstable and vary over time, cointegration measures long-term co-movements, being there even through sub-periods where correlation appears low.

The fourth measure used to pre-select pairs of stocks is the absolute value of the Pearson correlation coefficient between the "homoskedastic" returns, used as criterion of pairs formation by Chen et al. (2017), that is:

$$\hat{\rho}^r = \left| \frac{\sum_{t=1}^{T} \left({}_hr_{1,t} - {}_h\bar{r}_1\right)\left({}_hr_{2,t} - {}_h\bar{r}_2\right)}{\sqrt{\sum_{t=1}^{T} \left({}_hr_{1,t} - {}_h\bar{r}_1\right)^2 \sum_{t=1}^{T} \left({}_hr_{2,t} - {}_h\bar{r}_2\right)^2}} \right| \tag{4}$$

where ${}_hr_{1,t}$ and ${}_hr_{2,t}$ are the "homoskedastic" returns on day t of stock 1 and 2, respectively obtained as difference of the stock "homoskedastic" log-prices (i.e., ${}_hr_{1,t} = {}_hp_{1,t} - {}_hp_{1,t-1}$ and ${}_hr_{2,t} = {}_hp_{2,t} - {}_hp_{2,t-1}$), ${}_h\bar{r}_1$ and ${}_h\bar{r}_2$ are their corresponding sample means over the formation period.

Both $\hat{\rho}^P_{1,2}$ and $\hat{\rho}^r_{1,2}$, i.e. the correlation between "homoskedastic" log-prices and returns, tend to their maximum value as the standard deviation of the underlying series tend to their minimum. This means that selecting pairs by maximizing the absolute value of the correlation may result in the selection of stocks whose prices or returns display low volatilities. In order to overcome this drawback, as our fifth and sixth measures of pre-selection we consider the corresponding covariances between both "homoskedastic" log-prices and returns, that is:

$$\widehat{COV}^p = \sum_{t=1}^{T} \left[{}_hp_{1,t} - {}_h\bar{p}_1\right]\left[{}_hp_{2,t} - {}_h\bar{p}_2\right] \tag{5}$$

$$\widehat{COV}^r = \sum_{t=1}^{T} \left({}_hr_{1,t} - {}_h\bar{r}_1\right)\left({}_hr_{2,t} - {}_h\bar{r}_2\right) \tag{6}$$

All the measures considered so far are intuitive and computationally non-demanding, making them particularly suitable for practitioners undertaking high-frequency trading, where speed and computational efficiency are pivotal (Clark 2012; Brogaard et al. 2014; Angel 2014). However, they are not necessarily connected with the existence of a common trend between the paired assets. The seventh measure we consider aims to overcome this concern. The magnitude-squared coherence is a signal processing tool that indicates how well two signals match at each frequency and

[5] It is well known that the use of correlation measures between integrated processes is highly problematic, due to the fact that non-stationary processes are not ergodic. As proven by the simulations in Granger and Newbold (1974) and by the formal proofs in Phillips (1986), the R^2 of a regression between non-stationary processes that are not cointegrated do not converge in probability to a fixed value bur rather it has a non-degenerate asymptotic distribution.

Journal : SmallExtended 10260 | Article No : 702 | Pages : 32 | MS Code : 702 | Dispatch : 21-4-2023

its estimate is a function with values between 0 and 1. It measures the linear dependence in the spectral decomposition of $\Delta_h p_{1,t}$ and $\Delta_h p_{2,t}$ by computing $\widehat{C}_{\Delta_h p_{1,t}\Delta_h p_{2,t}}(f)$ values at different frequencies f as:

$$\widehat{C}_{\Delta_h p_{1,t}\Delta_h p_{2,t}}(f) = \frac{\left|\widehat{S}_{\Delta_h p_{1,t}\Delta_h p_{2,t}}(f)\right|^2}{\widehat{S}_{\Delta_h p_{1,t}\Delta_h p_{1,t}}(f)\widehat{S}_{\Delta_h p_{2,t}\Delta_h p_{2,t}}(f)} \tag{7}$$

where $\widehat{S}_{\Delta_h p_{1,t}\Delta_h p_{2,t}}(f)$ is the cross-power spectral density of $\Delta_h p_{1,t}$ and $\Delta_h p_{2,t}$, and $\widehat{S}_{\Delta_h p_{1,t}\Delta_h p_{1,t}}(f)$ and $\widehat{S}_{\Delta_h p_{2,t}\Delta_h p_{2,t}}(f)$ are the related power spectral densities.[6] In particular, we consider the magnitude-squared coherence at frequency zero, $\widehat{C}_{\Delta \mathrm{p}}(0)$ which is equal to 1 if the two series are cointegrated (Cubadda 1994; Levy 2002).

Each metrics is computed monthly using data from the previous one-year formation period. Stock pairs are then sorted in ascending order according to SSD, by PR closest to one, and in descending order according to measures (3) to (7). This first step thus closes with the pairs ranked on the basis of the metrics illustrated.

3.2 Cointegration-based pairs identification

In the second step, the ranked pre-selected pairs are progressively tested for cointegration until 20 pairs of stocks whose prices are actually cointegrated are found. Those "top 20" pairs are thus eligible for trading during the following six months.[7]

The most widely used procedure in the empirical literature on cointegration-based pairs trading is the two-step approach proposed by Engle and Granger (1987). Along with its simplicity, which makes it highly appreciated by practitioners, this procedure comes with several limitations, which include the impossibility to retrieve more than one cointegration relationship as well as the sensitivity of the results to the choice of the asset used as dependent variable. While the former is not an issue in pairs trading, as dealing with two assets at a time entails that at most one cointegration vector can be found, the second limitation might indeed be a serious issue. We thus opt for the Johansen (1988) procedure, where the test for the existence of cointegration and the estimates of the coefficient rely on a vector autoregression (VAR) model.[8] More specifically, consider the following VAR of order q:

$$Y_t = \mu + A_1 Y_{t-1} + \cdots + A_q Y_{t-q} + \varepsilon_t \tag{8}$$

where Y_t is the $T \times 2$ vector of the non-stationary, or $I(1)$, time-series of the log-prices of the two stocks, denoted with $p_{1,t}$ and $p_{2,t}$, ε_t is the $T \times 1$ vector of

[6] The power and cross-power spectral densities between $\Delta p_{1,t}$ and $\Delta p_{2,t}$ are computed using the Welch's averaged modified periodogram.

[7] As robustness check, we also considered the case of selecting the first 50 and 100 pairs, obtaining similar results, available upon request.

[8] This procedure has been proved to perform well even in presence of heteroskedasticity, which characterize financial time series, see Cavaliere et al. (2015) and Cavaliere et al. (2018), among others.

innovations, and q, the lag order, is defined optimally through the Bayesian Information Criterion (BIC). This VAR can be rewritten as:

$$\Delta Y_t = \mu + \Pi Y_{t-1} + \sum_{i=0}^{q-1} \Gamma_i \Delta Y_{t-i} + \varepsilon_t \tag{9}$$

with

$$\Pi = \sum_{i=0}^{q} A_i - I and \Gamma_i = - \sum_{j=i+1}^{q} A_j \tag{10}$$

The matrix Π can be then expressed in terms of the loadings vector α and the cointegration vector B as:

$$\Pi = \alpha \mathrm{B}' \tag{11}$$

The Johansen test for cointegration amounts to test for the rank of matrix Π, by testing whether (and how many of) its two eigenvalues are significantly different from zero. If they are both different from (or equal to) 0, then Π has full (zero) rank and there is no cointegration between the series. If instead Π has rank equal to 1, i.e. only one eigenvalue is statistically different from zero, then $\mathrm{B}' Y_t$ is stationary, implying that the time-series of the log-prices are cointegrated with cointegration vector B, and the two assets are thus considered eligible for trading.[9]

This concludes the second step of the procedure, which—as the first one—is repeated monthly in a rolling-window setting using data from the previous one-year formation period. This length for the formation period is almost a standard in the relevant literature, as it allows consistent estimates of the cointegration vector and takes up the potential calendar effects, pretty pronounced on the stock market (see e.g. Sharma and Narayan 2014, McConnell and Xu 2008, and Kunkel et al. 2003 and references therein).

We are now able to evaluate the consequences of ordering and pre-selecting pairs of assets before testing for cointegration in a pairs trading setting. The fewer the number of tests needed, the less time will be necessary to select pairs. As reported in the top panel of Table 1, pre-selecting assets allows to substantially reduce the total number of cointegration tests actually required to find the top 20 pairs of assets eligible for trading. The computational gain is thus remarkable, and this applies to all the pre-selection measures considered. On the other hand, the top 20 pairs eventually selected for trading are in most of the cases different depending on the pre-selection measure used (see the

[9] In our implementation we use the trace test, assume a model with intercepts in the cointegration vectors and deterministic linear trends in the levels of the data, and set the optimal lags based on the Bayesian Information Criterion.

We also implement the Gregory and Hansen (1996) test in the cases in which the standard cointegration tests fails to reject the null hypothesis of no cointegration to spot potential structural break in the cointegration relationship, finding that they are quite rare (between 0.06% of the cases, when the pre-selection is performed via covariance of log-prices, to slightly more than 5.0%, when SSD is used instead). We are thus reassured about the robustness of our conclusions to one potential change in the cointegration vector within each (one-year) single formation period.

Table 1 Tests required to spot the top 20 pairs and share of overlapping top 20 pairs, by pre-selection measure

	$\widehat{SSD}$	$\widehat{PR}$	$\hat{\rho}^p$	$\hat{\rho}^r$	$\widehat{COV}^p$	$\widehat{COV}^r$	$\hat{C}_{\Delta p}(0)$
	Comparison of the cointegration tests required						
Potential pairs	100,646	100,646	100,646	100,639	100,646	100,639	100,639
Tested pairs	31	24	22	26	21	25	24
	Comparison of the top 20 tradable pairs selected						
$\widehat{SSD}$	1	0.01	0.13	0.13	0.00	0.08	0.03
$\widehat{PR}$		1	0.00	0.00	0.00	0.00	0.00
$\hat{\rho}^p$			1	0.13	0.08	0.10	0.04
$\hat{\rho}^r$				1	0.00	0.41	0.06
$\widehat{COV}^p$					1	0.01	0.00
$\widehat{COV}^r$						1	0.06
$\hat{C}_{\Delta p}(0)$							1

The table reports the (average across formation periods) number of cointegration tests required to spot the top 20 pairs eligible for trading, along with the (average across formation periods) potential pairs to be tested, by pre-selection measure. $\widehat{SSD}$ is the sum of squared deviations between normalized log-prices, $\widehat{PR}$ is the price ratio of normalized log-prices, $\hat{\rho}^p$ is the log-prices correlation, $\hat{\rho}^r$ is the returns correlation, $\widehat{COV}^p$ is the log-prices covariance, $\widehat{COV}^r$ is the returns covariance, and $\hat{C}_{\Delta p}(0)$ is the magnitude-squared coherence between the first-difference of log-prices at frequency zero

bottom panel of Table 1). The maximum degree of overlap is indeed found between returns correlation and returns covariance, where the pre-selected pairs coincide in the 41% of the cases. For all the other combinations, the pairs selected overlap by at most 13%, with several cases in which the overlap is negligible or even non-existent. This confirms the potentially crucial impact of the pre-selection metrics used on the pairs that are actually traded and, hence, on the final profitability of the pairs trading strategy.

3.3 Pairs trading strategy implementation

Next, we move to the following six-month period, the so-called trading period, as it is where we implement the pairs trading strategy. As argued in Huck (2013), a six-month length represents a good choice as it ensures the trades to be performed on pairs selected based on the most recent information, on the one hand, and allows the reversion of the spread, and hence the trades to close "naturally", on the other. Thus, using data on the six-month following the trading period, we implement the self-financing pairs trading strategy whenever pricing anomalies are signaled by deviations from the long-run equilibrium relationship.

In order to do so, we compute the *Spread* between log-prices, which is stationary and mean-reverting, as:

Journal : SmallExtended 10260 | Article No : 702 | Pages : 32 | MS Code : 702 | Dispatch : 21-4-2023

$$Spread_t = p_{1,t} - \left(\hat{\mu} + \hat{\beta} p_{2,t}\right) \tag{12}$$

using the $\hat{\mu}$ and $\hat{\beta}$ estimates obtained in the first step of the procedure.[10] Any significant deviation of the *Spread* from its historical mean (equal to zero by construction) is interpreted as a mispricing and thus signals a trading opportunity. Therefore, we can then define a trading rule that will trigger a trade whenever the following relationship is violated:

$$-2\hat{\sigma} \leq Spread_t \leq 2\hat{\sigma} \tag{13}$$

where $\hat{\sigma}$ is the historical standard deviation of the *Spread* (computed during the formation period).

In more detail, if $Spread_t > 2\hat{\sigma}$ the first stock is suspected to be relatively overpriced with respect to the second. Then, the self-financing strategy consists in selling (going short on) 1\$ of the first stock and buying (going long on) 1\$ of the second stock. On the other hand, if $Spread_t < -2\hat{\sigma}$ the first stock is suspected to be underpriced relatively to the second stock. Therefore, the trading strategy requires to sell 1\$ of the second stock and to buy 1\$ of the first stock. Notice that the strategy is self-financing, since it requires no initial capital, and consequently all payoffs can be interpreted as excess returns. The position is unwound when the long-term equilibrium is re-established, that is, when the *Spread* returns to within the estimated boundaries[11] (or at the end of the trading period). An example of when a trading opportunity arises (green arrow) and when the position is closed (red arrow) is provided in Fig. 1.

This whole procedure, from step 1 to step 3, is then repeated in a rolling window fashion, by shifting the formation and trading periods by one month. As a result, every month (starting from the 6th in the sample) six overlapping portfolios of 20 pairs are generated (Fig. 2). This approach mimics the payoffs that a proprietary trading desk would get delegating the management of these portfolios to six different traders whose formation and trading periods are staggered by one month (see, e.g. Gatev et al. 2006 and Huck and Afawubo 2015).

Our dataset comprises 235 trading months.[12] Excluding the first and the last 5 months, which by construction do not include the full set of six overlapping portfolios, leaves us with 225 trading months, which are the ones used to evaluate the performance of the strategy.

[10] The parameters used are the ones belonging to the cointegrating vector $B' = [\mu\beta]$ estimated in the formation period, so that the Spread is the series of out-of-sample residuals.

[11] As robustness check, in Sect. 5.2, we also examine the case in which a position is closed whenever the Spread reaches zero.

[12] The dataset spans from January 1998 until October 2018 and comprises 5434 daily observations. Since a month includes an average of 22 trading days, and since the first 12 months are used for the first formation period, we end up with 235 trading months.

Journal : SmallExtended 10260 | Article No : 702 | Pages : 32 | MS Code : 702 | Dispatch : 21-4-2023

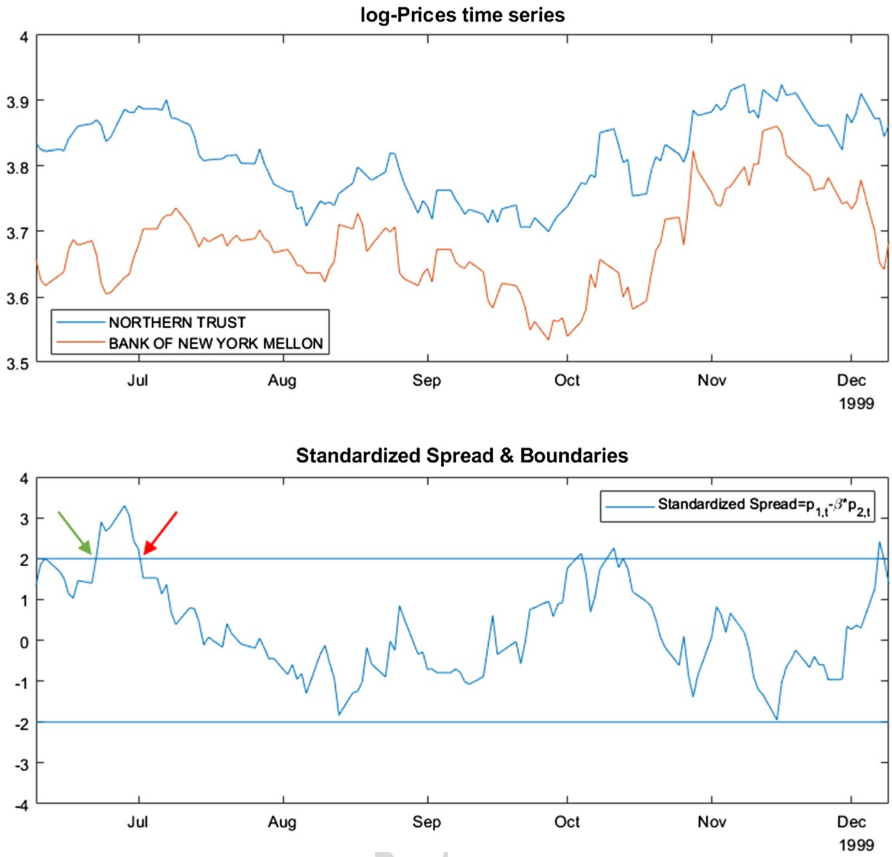

Fig. 1 - Example of trading strategy triggering signals. The top panel plots the log-prices time series of two cointegrated stocks, during the trading period going from 9th June to 9th December 1999. The bottom panel plots the standardized Spread along with its $\pm 2\hat{\sigma}$ boundaries. The green arrow spots the opening of the trading position, while the red one spots the closing of the same trade

3.4 Performance evaluation

The profitability of the pairs trading strategy is evaluated in terms of excess returns (profits), Sharpe ratio and percentage of positive monthly excess returns.

The value-weighted mark-to-market daily excess return of the portfolio constituted by the 20 traded pairs is computed as:

$$R_{P,t} = \frac{\sum_{i=1}^{20} w_{i,t} R_{i,t}}{\sum_{i=1}^{20} w_{i,t}} \tag{14}$$

where:

Journal : SmallExtended 10260	Article No : 702	Pages : 32	MS Code : 702	Dispatch : 21-4-2023

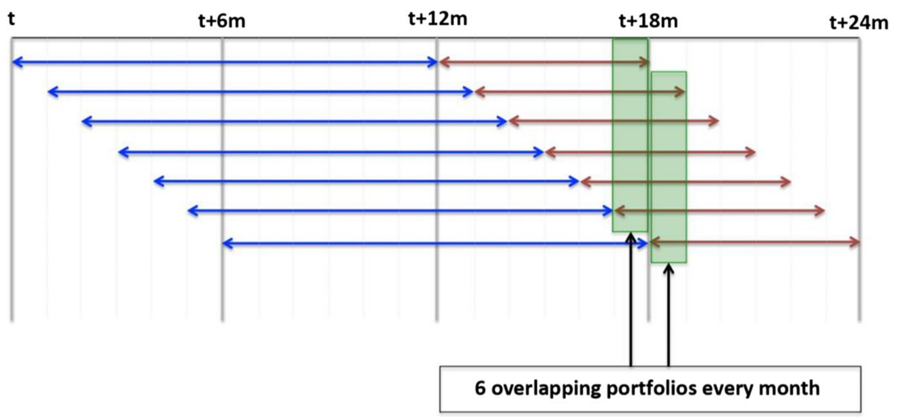

Fig. 2 - Rolling scheme representation. Graphical representation of the rolling scheme employed in the pairs trading methodology. The blue arrows represent the length of the formation periods, the brown arrows represent the length of the trading periods and the green areas represent the overlapping months across different trading periods. The time indicator "t" refers to the t-th month (composed of 22 trading days)

- $w_{i,t}$ is the weight associated to each pair i. It is equal to 1 whenever a new position is opened on the pair i and, for each subsequent period, is computed as: $w_{i,t} = w_{i,t-1}\left(1 + R_{i,t-1}\right) = \left(1 + R_{i,1}\right) \dots \left(1 + R_{i,t-1}\right)$with:
- $R_{i,t}$ is the daily mark-to-market excess return obtained from trading pair i, computed as:with:

$$R_{i,t} = \sum_{j=1}^{2} I_{j,t}\omega_{j,t}R_{j,t} \tag{15}$$

- $I_{j,t}$ denoting the position on stock j in day t. It is set equal to 1 if a long position is open, -1 if a short position is opened, and 0 otherwise
- $R_{j,t}$ being the daily return of stock j in day t
- $\omega_{j,t}$ representing the weight associated to stock $j$inday$t$, computed as $\omega_{j,t} = \omega_{j,t-1}\left(1 + R_{j,t-1}\right) = \left(1 + R_{j,1}\right) \dots \left(1 + R_{j,t-1}\right)$

The daily portfolio excess returns are then compounded to obtain monthly excess returns, which, in turn, are averaged across the six overlapping portfolios so as to generate a single summarizing monthly measure of profitability.

Finally, the null that the average monthly excess returns are not significantly positive is tested using Newey-West (1987) heteroskedasticity and autocorrelation robust standard errors, using six lags. Following Huck and Afawubo (2015), we handle the potential data-snooping issue deriving from testing multiple strategies on the same data set by using the Hansen (2005) test for Superior Predictive Ability, which takes into account the dependence between the statistics derived from the different implementations.

Table 2 Commissions' estimates from the quarterly reports of the ITG

Year	Institutional trades average commissions (bps)
1998	10
1999	10
2000	10
2001	10
2002	10
2003	10
2004	10
2005	10
2006	9
2007	7
2008	8
2009	9
2010	8
2011	7
2012	7
2013	6
2014	5
2015	5
2016	5
2017	4
2018	3

Besides the excess returns, we also evaluate the profitability of the pairs trading strategy based on the Sharpe ratio and on the percentage of positive monthly excess returns. The first one accounts also for the volatility of profits, as it rescales the average monthly excess return by the monthly excess returns' standard deviation. The second is particularly suitable considered that pairs trading is a statistical—rather than a pure—arbitrage strategy, which thus allows negative returns. An account of the frequency of positive monthly excess returns is thus certainly relevant.

Transaction costs might have a critical impact on the profitability of a trading strategy and have thus to be carefully evaluated. The most well-known components of direct trading costs include commissions, short selling fees and bid and ask spread. In Gatev, et al. (2006) the bid-ask spread is handled by waiting one day after the divergence (convergence) to open (close) a position. Yet, the data used for the empirical application in this paper refer to assets with an extremely high level of liquidity, which are generally traded over a relatively short period, and have high dollar value and high market capitalization. Hence, bid and ask spread is likely not to be an issue. The same applies to short-selling, as reported in D'Avolio's (2002) and referred to also in Do and Faff (2012). Based on this claims, in this study the final profits from pairs trading are not weighted up with short-selling fees and bid-ask spread. On the other hand, commissions are likely to be particularly influential for pairs trading, where two roundtrip transactions are involved. Following Do and Faff (2012), we take commissions estimates from reports of the Investment Technology Group (ITG), a brokerage firm specialized in trade execution. The values of the commissions used are reported, in basis points, in Table 2.[13]

One of the limitations posed by including commissions is that the trading strategy can no longer be considered self-financing. Subtracting the commissions amount from the cash flows generated by the position opening, would indeed require an initial capital equal to the total commissions for the two stocks. To overcome this limitation and to obtain results that can still be interpreted as excess returns, we adapt the amounts of each stock that are bought or sold. More specifically, when opening a position the amount (1-commission)$ is bought and the amount (1 + commission)$ is sold, so that the total initial cash flow remains zero and the strategy remains self-financing. When closing the position, the commissions paid are included by adjusting the daily excess returns for the pairs, as follows:

$$R_{i,t} = \sum_{j=1}^{2} \left(I_{j,t}\omega_{j,t}R_{j,t} - c\omega_{j,t}\left(1 + R_{j,t}\right)\right) \quad (16)$$

where c is the amount of the commission as a percentage.

Besides, following Caldeira and Moura (2013), we complete our trading strategy with a stop-loss rule, which prescribes to close a position whenever the realized excess return of the operation reaches -7%, and a “duration rule”, which specifies to forcibly close a position after 50 trading days. These cut rules handle the risks of extreme losses, as well as the possibility that the *Spread* does not revert to its equilibrium. While academic research might disregard these aspects, practitioners are strongly aimed at preventing extremely negative results and loosing time value.

In this work, we thus evaluate the profitability both with and without taking transaction commissions and cut rules into account.[14]

4 Main results

The top panel of Table 3 reports the descriptive statistics of the monthly excess returns obtained from the pairs trading strategy when neither commissions nor cut rules are considered. The profitability of the strategy remarkably differs depending on the pre-selection metrics used. When pre-selection is carried out by means of

[13] Data from 1998 to 2009 are directly taken from Do and Faff (2012), while for the following years we compute average annual commissions from quarterly data published in the fourth quarter of 2018 ITG report.

[14] Besides direct costs, one might also consider the implicit cost represented by the market impact. This is evaluated separately in Sect. 5.3, not only because it is an implicit rather than a direct cost, but also because once these costs are taken into account the strategy is no longer self-financing and the final profits can no longer be interpreted as excess returns.

Table 3 Pairs trading monthly excess returns statistics, by pre-selection metrics

Pre-selection metrics	$\widehat{SSD}$	$\widehat{PR}$	$\hat{\rho}^p$	$\hat{\rho}^r$	$\widehat{COV}^p$	$\widehat{COV}^r$	$\hat{C}_{\Delta p}(0)$
Before commissions and cut rules (in \$)							
Mean	0.0033	0.0070	0.0073	0.0032	0.0076	0.0056	0.0019
Standard deviation	0.0235	0.0284	0.0369	0.0266	0.1082	0.0371	0.0319
Min | Max	– 0.06 | 0.09	– 0.06 | 0.14	– 0.15 | 0.14	– 0.09 | 0.11	– 0.36 | 0.91	– 0.08 | 0.27	– 0.10 | 0.10
Median	0.0013	0.0034	0.0058	0.0036	– 0.0015	0.0010	0.0003
NW t-statistics (*p* Value)	2.88 (0.00)	3.16 (0.00)	2.87 (0.00)	1.76 (0.04)	1.13 (0.13)	2.62 (0.00)	0.86 (0.20)
Consistent *p* Value	0.00	0.00	0.01	0.04	0.11	0.02	0.20
Sharpe ratio	0.14	0.25	0.20	0.12	0.07	0.15	0.06
% positive excess returns	51.91%	54.47%	54.47%	56.17%	46.81%	50.64%	48.94%
Average life (days)	24.17	19.71	23.90	18.56	23.49	17.89	20.74
Including commissions and cut rules (in \$)							
Mean	0.0003	0.0055	0.0108	0.0018	0.0067	0.0031	0.0013
Standard deviation	0.0225	0.0323	0.0438	0.0250	0.0952	0.0331	0.0307
Min | Max	– 0.07 | 0.09	– 0.08 | 0.25	– 0.11 | 0.24	– 0.07 | 0.10	– 0.39 | 0.61	– 0.08 | 0.24	– 0.12 | 0.10
Median	– 0.0004	0.0035	0.0030	0.0015	0.0007	– 0.0008	– 0.0021
NW t-statistics (*p* Value)	0.21 (0.42)	2.05 (0.02)	2.97 (0.00)	1.05 (0.15)	1.20 (0.11)	1.54 (0.06)	0.59 (0.28)
Consistent *p* Value	0.44	0.02	0.01	0.14	0.13	0.08	0.26
Sharpe ratio	0.01	0.17	0.25	0.07	0.07	0.10	0.04
% positive excess returns	46.81%	52.34%	51.49%	50.21%	49.36%	46.38%	43.40%
Average life (days)	13.55	11.15	12.40	11.95	9.78	11.39	11.57

The table reports the main descriptive statistics of the excess returns obtained implementing a cointegration-based pairs trading strategy when pairs are pre-selected using different metrics, namely: sum of squared deviations between normalized log-prices, $\widehat{SSD}$, price ratio of normalized log-prices, $\widehat{PR}$, log-prices correlation $\hat{\rho}^p$, returns correlation $\hat{\rho}^r$, log-prices covariance $\widehat{COV}^p$, returns covariance $\widehat{COV}^r$, and the magnitude-squared coherence between the first-difference of log prices at frequency zero $\hat{C}_{\Delta p}(0)$. The null that the average monthly excess returns are not significantly positive is tested using Newey-West (1987) heteroskedasticity and autocorrelation robust t-statistics (associated *p* Values in parenthesis) and the Hansen (2005) consistent *p* Value, to control for the risk of data-snooping the log-prices covariance or by the coherence metrics, the pairs trading strategy is found to produce non-significant average excess returns. All other pre-selection metrics instead lead to significant profitability. Moreover, the average excess returns display quite a large degree of variability also in terms of level: the (significant) 0.73% obtained when pairs are pre-selected based on the log-prices correlation, is more than 4 times the (non-significant) 0.19% average net profit obtained carrying out the very same trading strategy but using the coherence metrics to pre-select instead.

The picture does not change if the median—rather than the mean—excess return is considered: the highest median excess return, obtained pre-selecting pairs based on correlation between log-prices (0.58%), is almost 6 times the lowest one (− 0.15%), obtained using the covariance between returns as pre-selection measure.

The profitability observed for the different pre-selection measures is highly heterogeneous also in terms of variability, with SSD and covariance being at the extremes. In fact, the excess returns obtained with pairs pre-selected using SSD have a standard deviation (range) equal to 0.0235 (0.15), much smaller than the one observed when pre-selection occurs based on covariance of log-prices, equal to 0.1082 (1.27).

Not surprisingly, the profitability per unit of risk, as measured by the Sharpe Ratio, is not uniform across the different pre-selection metrics considered: the Sharpe Ratio is around 0.06 when pre-selection is based on the covariance of log-prices or on spectral coherence, it almost doubles when returns correlation or SSD are used to pre-select, and it further increases, reaching as much as 0.25, when pre-selection is done via PR.

The effect of using different pre-selection measures is also evident looking at the share of positive monthly excess returns. When pairs are pre-selected using the log-prices covariance and coherence this share is less than half, while it increases to around 54% if pre-selection uses log-prices correlation or Price Ratio and to 56.17% when pairs are pre-selected based on returns correlation.

To sum up, we find that when the same cointegration-based pairs trading strategy is implemented pre-selecting pairs with different metrics, the final profitability obtained is highly heterogenous in terms of average, variability and significance. We thus conclude that pre-selection matters, as it might lead to remarkably different final excess returns.

The impact of pre-selecting is even more apparent when the profitability is evaluated taking commissions and cut rules into account (see Panel B, Table 3). Indeed, the average monthly excess returns are not statistically significant in 4 cases out of 7 (namely when pre-selection occurs via SSD, returns correlation, covariance of log-prices and coherence). Besides, when statistically significant, the excess returns appear to be remarkably different, ranging from 0.31%, obtained when pairs are pre-selected based on the returns covariance, to 1.08%, as much as 3 times more, when pre-selection is performed using the log-prices correlation. Additionally, the Sharpe Ratios vary by a factor of 25 among all the metrics, ranging from 0.01, when pre-selection is based on SSD, to as much as 0.25, when pairs are pre-selected via log-prices correlation. The impact of pre-selection is also confirmed looking at the frequency of positive monthly excess returns: correlation between log-prices and PR

Journal : SmallExtended 10260 | Article No : 702 | Pages : 32 | MS Code : 702 | Dispatch : 21-4-2023

guarantee the highest frequencies (51.49% and 52.34%, respectively), while all others pre-selection metrics reduce this chance around or well below 50%.

Finally, a comparison between the top and bottom panels of Table 3 allows to assess the differential impact of commissions and cut rules across the different pre-selection metrics. Recall that the commissions and the cut rules have opposite effects on profitability, so that the overall final effect might be either positive, null or negative, depending on which one prevails, if any. Again, we find disparate results across the different metrics considered: e.g., for SSD the introduction of commissions and cut rules translates into remarkably lower average excess returns, for log-prices correlation it leads to better results, while for others, such as coherence, the overall impact is found to be negligible. By the same token, the rankings based e.g. on Sharpe Ratio or frequency of positive excess returns—before and after commissions and cut rules are taken into account—change for some pre-selection metrics more than others, further confirming that the pre-selection measure used have an impact also in this direction.

All in all, the evidence reported proves that pre-selection matters, since it strongly impacts the final profitability of the pairs trading strategy. Remarkably, none among the three additional pre-selection investigated seem to produce systematically better results compared to ones traditionally considered in the pairs trading literature.

5 Further results

In this section, we first investigate whether the monthly excess returns observed represent a compensation for traditional risk factors. Then, we check the robustness of our main conclusions to a stricter definition of the Spread reversion to the equilibrium, to the inclusion of market impact on the final evaluation of the strategy profitability, and to the use of Engle and Granger (1987) procedure to assess cointegration.

5.1 Excess returns and risk factors

We investigate if and how the pairs trading profitability obtained applying each of the considered pre-selection metrics correlates with the systematic stock-market risk factors conventionally acknowledged in the asset pricing literature. To this end, we regress the monthly excess returns (after commision and cut rules are included) on the following factors[15]:

1. Market excess return (MKT): difference between the market (S&P500) and 30-day Treasury bill returns. This factor is key to assess market-neutrality: to the extent that pairs trading is a market-neutral strategy, the correlation between its profitability and the market excess returns is expected to be small, if any.
2. Size factor (SMB): difference between small and big stock portfolios;

15FL01 15FL02 15FL03 [15] Data and more detailed description are available on Kenneth R. French website: https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html. The website provides daily data, which we compound in order to obtain monthly values.

3. Book-to-market factor (HML): difference between value and growth stock portfolios.

These first three are the stock-market risk factors established in the seminal paper by Fama and French (1993).

4. Investment factor (CMA): difference between conservative and aggressive portfolios;
5. Profitability factor (RMW): difference between robust and weak profitability portfolios.

This two additional factors lead to the five-factor model proposed in Fama and French (2015). Since pairs trading is a contrarian strategy, we expect a mild correlation, if any, of its profitability to these factors.

6. Short-term reversal factor (STR): difference between last month winner and loser portfolios. Recalling that trading on assets which performed particularly well in the past month has been proven to lead to significant abnormal returns (see, e.g. Jegadeesh 1990, Lehmann 1990, or Jegadeesh and Titman 1993) and that pairs trading sells overpriced (i.e. winners) and buys underpriced (losers) assets, the excess returns of pairs trading are expected to be positively exposed to this factor.
7. Momentum factor (MOM): difference between last year winner and loser portfolios; to the extent that pairs trading is a short-term strategy, we expect its excess returns to correlate negatively with the momentum factor.

These last two factors are included in the light of the fact that pairs trading is, by construction, a trading strategy relying on the past (patterns of) prices/returns of the assets involved.

Various combinations of the above-mentioned factors have been extensively used in the literature to assess whether the observed price anomalies, and hence the associated arbitrage opportunities, are priced by the common risk-factors (see e.g. Heiko 2015, and reference therein). The pairs trading literature we refer to makes no exception. For instance, Gatev et al. (2006) and Engelberg et al. (2009) employ the standard Fama and French (1993) three-factors model augmented by momentum and short term reversal factors, while Rad et al. (2016) use both the Fama and French (1993) three-factors model augmented with momentum and liquidity factors and the Fama and French (2015) five-factors model. More recently, both Clegg and Krauss (2018) and Han et al. (2022) regress the excess returns of the pairs trading strategy on the Fama and French (1993) three factors, Fama and French (2015) five factors, and Fama and French (1993) three factors augmented with momentum and short-term reversal factors.

Following Rad et al. (2016), we opt for the most comprehensive model specification, including all the factors at a time, because we also want to evaluate the intercept of the estimated model. Indeed, the intercept of this model can be interpreted as the alpha in a CAPM, whose value captures the average excess return of

Journal : SmallExtended 10260 | Article No : 702 | Pages : 32 | MS Code : 702 | Dispatch : 21-4-2023

Table 4 Pairs trading monthly excess returns risk-profile, by pre-selection metrics

	$\widehat{SSD}$	$\widehat{PR}$	$\widehat{\rho}^p$	$\widehat{\rho}^r$	$\widehat{COV}^p$	$\widehat{COV}^r$	$\widehat{C}_{\Delta p}(0)$
Intercept	0.001	0.004**	0.003	− 0.0002	0.011	0.003	− 0.0003
	(0.002)	(0.002)	(0.002)	(0.002)	(0.007)	(0.003)	(0.002)
Mkt	0.082**	0.055	0.185***	0.065	− 0.135	0.047	0.047
	(0.034)	(0.042)	(0.054)	(0.041)	(0.153)	(0.058)	(0.048)
SMB	− 0.059	0.268***	− 0.038	0.005	0.289	0.106	0.027
	(0.058)	(0.071)	(0.093)	(0.07)	(0.263)	(0.099)	(0.083)
HML	0.07	− 0.138*	0.196**	0.233***	0.725***	− 0.241**	0.105
	(0.06)	(0.074)	(0.096)	(0.073)	(0.273)	(0.102)	(0.086)
CMA	0.054	0.049	0.193	− 0.018	0.413	0.335**	− 0.044
	(0.092)	(0.113)	(0.146)	(0.111)	(0.416)	(0.156)	(0.131)
RMW	− 0.085	0.168**	− 0.193*	− 0.056	− 0.861***	0.232**	− 0.285***
	(0.068)	(0.083)	(0.108)	(0.082)	(0.308)	(0.115)	(0.097)
STR	0.008	− 0.041	0.037	− 0.019	− 0.233	− 0.101*	0.038
	(0.034)	(0.042)	(0.055)	(0.042)	(0.155)	(0.058)	(0.049)
MOM	− 0.123***	− 0.196***	− 0.065	0.004	− 0.692***	− 0.173***	− 0.059
	(0.031)	(0.038)	(0.05)	(0.038)	(0.141)	(0.053)	(0.044)
Obs	225	225	225	225	225	225	225
R^2	0.201	0.175	0.176	0.090	0.259	0.080	0.129
Adjusted R^2	0.176	0.149	0.150	0.060	0.235	0.051	0.101

The table reports the estimates obtained regressing monthly excess returns (after commissions and cut rules) against the following factors: Market Excess Return (MKT), Small minus Big (SMB), High minus Low (HML), Conservative minus Aggressive (CMA) and Robust minus Weak (RMW), Short-term Reversal (STR), and Momentum (MOM). The intercept of the regression captures the excess return achieved above the expected one based on the risk factors included in the regression. Each column reports the results referred to monthly excess returns obtained using different pre-selection metrics, namely: the sum of squared deviations between normalized log-prices, $\widehat{SSD}$, the price ratio of normalized log-prices, $\widehat{PR}$, the log-prices correlation $\widehat{\rho}^p$, the returns correlation $\widehat{\rho}^r$, the log-prices covariance $\widehat{COV}^p$, the returns covariance $\widehat{COV}^r$, and the magnitude-squared coherence between the first-difference of log prices at frequency zero $\widehat{C}_{\Delta p}(0)$. Robust standard errors are given in parentheses

$*$ significant at 10% level. $**$ significant at 5% level. $***$ significant at 1% level

this strategy with respect to the one achieved by the market based on the risk factors included in the regression. A significant and positive alpha, even after controlling for all the possible common risk-factors, would imply that carrying out such an active strategy is able to beat the market.

As reported in Table 4, the results are highly disparate with reference to all the risk factors considered, thus proving that pre-selection impacts also on the risk-profile of the pairs trading eventual profitability. More specifically, the trading excess returns correlate with the market excess returns only if pairs are pre-selected via SSD or correlation between log-prices: hence, in these cases only the market-neutrality of the pairs trading strategy would be disproved. By constrast, when pre-selction is based on any other of the remaining metrics considered, the evidence would be in support of market neutrality. Likewise, we find that excess returns of pairs trading are positively correlated with size factor only when pre-selection occurs via

Price Ratio, while in all other cases no significant relationship is retreived. Different pre-selection metrics also translates into different degrees of correlation with the Book-to-market risk-factor, which is strongly related to excess returns only when pairs are pre-selected via correlation or covariance measures. Besides, we observe that covariance between returns is the only pre-selection metrics leading to profits positively correlated to both the CMA and RMW factors. In all other cases, we observe no correlation with CMA and in the majority of the cases a negative one, if any, with RMW. In the latter case, the estimates also show a remarkable variabilty in terms of magnitude, ranging from − 0.285 to 0.168. Finally, no significant correlation is found with the short-term reversal (STR), while the pairs trading excess returns often negatively correlate with the momentum factor (MOM). This result is consistent with the expectations and with the previous literature but again is far from being homogeous across all the metrics used for pre-selection, being observed in only 4 cases out of 7 (i.e. when pre-selection is carried out via SSD, PR and covariances measures).

Last, we find non-homogeneous results across the metrics also in terms of the model intercept. The set of risk factors considered is generally able to sweep away the significance of the intercept, leading to the conclusion that pairs trading do not sistematically produces extra-profits with respect to the market. However, this does not happen when pre-selection runs based on Price Ratio, where the positive and significant alpha suggests instead that implementing a pairs trading strategy pre-selecting asssets with this measure might actually beat the market.

All in all, this evidence proves that pre-selection matters also in terms of if and how the pairs trading final profitability relates to the conventional risk-factors.

5.2 Spread reverting to zero

We now evaluate the robustness of our results adopting an alternative definition of the reversion to the equilibrium that is stricter than the one used in the baseline approach. More specifically, we now close the positions whenever the Spread reaches zero (or at the end of the trading period), rather than just reentering within the $\pm 2\hat{\sigma}$ boundaries. The results, reported in Table 5 confirm the high degree of variation across the pre-selection metrics. For instance, after the inclusion of commissions and cut rules, the average excess returns range from values not statistically distinguishable from 0 (when using SSD, PR and coherence as pre-selection metrics) to as much as 1.26%, when pre-selection occurs with log-prices covariance. Similarly, the Sharpe Ratios range from almost 0, when using the SSD, up to 0.25, when pairs are pre-selected based again on log-prices correlation. Finally, the frequency of positive monthly excess returns varies of almost 12 percentage points, moving from above 56% when pre-selection is based on returns correlation, to 44% when pairs are pre-selected based on spectral coherence.

In general, imposing this stricter condition to unwound the positions on the assets implies a longer duration of each trade (from an average of around 12 days to an average of around 30 days, see bottom part of Tables 3 and 5), coupled with a higher chance of the 50-days cut rule kicking in. This means that more trades are forcibly

Table 5 Pairs trading monthly excess returns when spread reverts to 0, by pre-selection metrics

Pre-selection metrics	$\widehat{SSD}$	$\widehat{PR}$	$\hat{\rho}^p$	$\hat{\rho}^r$	$\widehat{COV}^p$	$\widehat{COV}^r$	$\widehat{C}_{\Delta p}(0)$
Before commissions and cut rules (in $)							
Mean	0.0004	0.0013	0.0053	0.0003	0.0042	0.0019	– 0.0012
Standard deviation	0.0215	0.0219	0.0385	0.0228	0.1131	0.0306	0.0272
Min ǀ Max	– 0.07 ǀ 0.09	– 0.07 ǀ 0.08	– 0.20 ǀ 0.28	– 0.11 ǀ 0.11	– 0.36 ǀ 0.96	– 0.14 ǀ 0.26	– 0.09 ǀ 0.10
Median	– 0.0010	0.0010	0.0022	0.0026	0.0011	0.0003	– 0.0014
NW t-statistics (*p* Value)	0.37 (0.35)	0.89 (0.19)	1.88 (0.03)	0.18 (0.43)	0.69 (0.24)	1.43 (0.08)	– 0.58 (0.72)
Consistent *p* Value	0.34	0.18	0.04	0.45	0.24	0.09	1.00
Sharpe ratio	0.02	0.06	0.14	0.01	0.04	0.06	– 0.04
% positive excess returns	44.26%	48.94%	50.64%	54.04%	48.09%	48.94%	43.40%
Average life (days)	66.64	60.82	63.77	60.74	64.39	58.44	61.60
Including commissions and cut rules (in $)							
Mean	0.0002	0.0009	0.0097	0.0021	0.0126	0.0035	0.0006
Standard deviation	0.0259	0.0294	0.0381	0.0253	0.0843	0.0331	0.0278
Min ǀ Max	– 0.07 ǀ 0.19	– 0.12 ǀ 0.11	– 0.15 ǀ 0.21	– 0.08 ǀ 0.17	– 0.29 ǀ 0.35	– 0.17 ǀ 0.20	– 0.15 ǀ 0.11
Median	– 0.0006	0.0008	0.0045	0.0027	0.0020	0.0022	– 0.0017
NW t-statistics (*p* Value)	0.14 (0.44)	0.43 (0.33)	3.57 (0.00)	1.68 (0.05)	2.19 (0.01)	1.91 (0.03)	0.30 (0.38)
Consistent *p* Value	0.44	0.32	0.00	0.06	0.01	0.06	0.41
Sharpe ratio	0.01	0.03	0.25	0.08	0.15	0.11	0.02
% positive excess returns	46.38%	49.79%	53.19%	56.17%	50.64%	54.47%	44.26%
Average life (days)	33.41	28.56	30.11	32.86	23.88	30.91	29.53

The table reports the main descriptive statistics of the excess returns obtained implementing a cointegration-based pairs trading strategy in which the trades are closed whenever the Spread reverts to zero, rather than just re-entering the $\pm 2\hat{\sigma}$ boundaries. The pairs are pre-selected using different metrics, namely: the sum of squared deviations between normalized log-prices, $\widehat{SSD}$, the price ratio of normalized log-prices, $\widehat{PR}$, the log-prices correlation $\hat{\rho}^p$, the returns correlation $\hat{\rho}^r$, the log-prices covariance $\widehat{COV}^p$, the returns covariance $\widehat{COV}^r$, and the magnitude-squared coherence between first-difference log prices at frequency zero $\widehat{C}_{\Delta p}(0)$. The null that the average monthly excess returns are not significantly positive is tested using Newey-West (1987) heteroskedasticity and autocorrelation robust t-statistics (associated *p* Values in parenthesis) and the Hansen (2005) consistent *p* Value, to control for the risk of data-snooping closed when the prices of the paired assets are still largely divergent, leading to more extreme outcomes in terms of final profitability. This might explain why, in some cases (e.g. when pre-selection is based on returns correlation and on covariance metrics) the average excess returns after commissions and cut rules are even higher than in the baseline case, and why the variation across the pre-selection measures is in this case generally higher compared to the baseline case.

5.3 Market impact

When big investors trade, market impact, i.e. the implicit costs entailed by the movement of (huge amount of) assets, has also to be taken into account as an additional contribution to transaction costs. Do and Faff (2012) estimate the average market impact for the US stock market equal to 26 basis points, if the sample period considered spans from 1963 to 2009, which reduces to 20 basis points over the sub-period going from 1989 to 2009. Since our sample covers the 1998–2008 period, we set the cost associated to market impact to 20 basis points of the traded amounts in dollars. For each transaction, we thus compute the market impact for the opening and closing days only in dollars. Then, for each day we compute the average amount of market impact across the traded pairs, and obtain the daily net profits as difference between the average daily excess returns minus the average amount of market impact as computed above. Daily net profits are, as usual, compounded to obtain monthly net profits, and then averaged across the six overlapping portfolios so as to generate the single summarizing monthly measure. Table 6 reports the main descriptive statistics of these monthly measures. For convenience, in the first row we also report the average monthly net profit before market impact is accounted for (and after the inclusion of commissions and cut rules, as previously reported in the bottom panel of Table 3Table 3). Notice that once market impact is considered, the results can no longer be interpreted as excess returns, since the strategy has now an initial cost of \$0.004 (that is the value of the market impact, fixed to 20 basis points, multiplied by \$1 for each side of the trade).

Once market impact is considered, monthly net profits suffer an average reduction of about \$0.005, and the pairs trading profits retaining statistical significance (albeit at 10% level only) are the ones obtained when pairs pre-selection is based on log-prices correlation only. In all the remaining cases final profits are not statistically distinguishable from 0, further confirming the relevant impact of using different metrics for pairs pre-selection on the eventual pairs trading profitability.

5.4 Engle and Granger (1987) cointegration test

In our main analysis, the potential cointegration relationship between pre-selected assets is estimated and tested via the Johansen (1988) procedure. In this section we replicate the analysis using the two-step Engle and Granger (1987) test to find the 20 pairs whose prices are actually cointegrated. Despite its limitations as a statistical tool, this procedure is computationally simple, and hence particularly suitable for

Journal : SmallExtended 10260 | Article No : 702 | Pages : 32 | MS Code : 702 | Dispatch : 21-4-2023

Table 6 Pairs trading profitability after commission and cut rules, by pre-selection metrics: the effects of market impact

Pre-selection metrics	$\widehat{SSD}$	$\widehat{PR}$	$\hat{\rho}^p$	$\hat{\rho}^r$	$\widehat{COV}^p$	$\widehat{COV}^r$	$\hat{C}_{\Delta p}(0)$
Average net profit before market impact	0.0003	0.0055	0.0108	0.0018	0.0067	0.0031	0.0013
Average net profit after market impact	− 0.0045	0.0006	0.0051	− 0.0030	0.0005	− 0.0018	− 0.0036
Standard deviation	0.0223	0.0318	0.0421	0.0244	0.0940	0.0321	0.0302
Min | Max	− 0.08 | 0.08	− 0.09 | 0.23	− 0.11 | 0.22	− 0.08 | 0.09	− 0.40 | 0.60	− 0.09 | 0.22	− 0.13 | 0.09
NW t-statistics (p Value)	-3.23 (1.00)	0.22 (0.41)	1.52 (0.06)	-1.83 (0.97)	0.10 (0.46)	− 0.95 (0.83)	-1.63 (0.95)
Consistent p Value	1.00	0.38	0.06	1.00	0.48	1.00	1.00
% positive profits	37.02%	45.11%	46.81%	42.55%	42.98%	36.60%	39.15%

The table reports the main descriptive statistics of the monthly net profits before(first line) and after (second line) market impact is taken into account obtained implementing a cointegration-based pairs trading strategy when pairs are pre-selected using different metrics, namely: sum of squared deviations between normalized log-prices,

$\widehat{SSD}$, price ratio of normalized log-prices, $\widehat{PR}$, log-prices correlation $\hat{\rho}^p$, returns correlation $\hat{\rho}^r$, log-prices covariance $\widehat{COV}^p$, returns covariance $\widehat{COV}^r$, and the magnitude-squared coherence between the first-difference of log prices at frequency zero $\hat{C}_{\Delta p}(0)$. The null that the average monthly excess returns are not significantly positive is tested using Newey-West (1987) heteroskedasticity and autocorrelation robust t-statistics (associated p Values in parenthesis) and the Hansen (2005) consistent p Value, to control for the risk of data-snooping

Table 7 Pairs trading excess returns by pre-selection metrics: Robustness using Engle-Granger procedure

Pre-selection metrics	$\widehat{SSD}$	$\widehat{PR}$	$\hat{\rho}^p$	$\hat{\rho}^r$	$\widehat{COV}^p$	$\widehat{COV}^r$	$\hat{C}_{\Delta p}(0)$
PANEL A							
Before commissions and cut rules (in \$)							
Mean	0.0072	0.0079	0.0100	0.0037	0.0093	0.0060	0.0075
Standard deviation	0.0285	0.0291	0.0336	0.0274	0.0724	0.0296	0.0276
Min ∣ Max	− 0.06 ∣ 0.21	− 0.09 ∣ 0.14	− 0.09 ∣ 0.14	− 0.13 ∣ 0.13	− 0.28 ∣ 0.60	− 0.12 ∣ 0.17	− 0.08 ∣ 0.12
Median	0.0021	0.0050	0.0077	0.0030	0.0026	0.0050	0.0049
NW t-statistics (*p* Value)	3.46 (0.00)	3.62 (0.00)	4.31 (0.00)	2.36 (0.01)	1.87 (0.03)	3.11 (0.00)	3.71 (0.00)
Consistent *p* Value	0.00	0.00	0.00	0.01	0.03	0.01	0.00
Sharpe ratio	0.25	0.27	0.30	0.14	0.13	0.20	0.27
% positive excess returns	50.21%	56.17%	61.28%	55.32%	51.91%	56.60%	58.72%
Including commissions and cut rules (in \$)							
Mean	0.0049	0.0072	0.0131	0.0027	0.0093	0.0054	0.0064
Standard deviation	0.0312	0.0308	0.0423	0.0249	0.0667	0.0299	0.0274
Min ∣ Max	− 0.10 ∣ 0.27	− 0.10 ∣ 0.17	− 0.09 ∣ 0.18	− 0.06 ∣ 0.11	− 0.16 ∣ 0.32	− 0.07 ∣ 0.20	− 0.08 ∣ 0.09
Median	− 0.0003	0.0044	0.0062	0.0012	0.0010	0.0021	0.0036
NW t-statistics (*p* Value)	2.14 (0.02)	3.19 (0.00)	3.62 (0.00)	1.90 (0.03)	2.14 (0.02)	2.28 (0.01)	2.83 (0.00)
Consistent *p* Value	0.04	0.00	0.00	0.04	0.02	0.01	0.00
Sharpe ratio	0.16	0.24	0.31	0.11	0.14	0.18	0.23
% positive excess returns	47.23%	54.04%	57.02%	50.64%	48.51%	53.19%	53.62%
PANEL B							
Before commissions and cut rules (in \$)							
Mean	0.0075	0.0064	0.0085	0.0054	0.0095	0.0067	0.0084
Standard deviation	0.0261	0.0302	0.0351	0.0288	0.0756	0.0258	0.0267
Min ∣ Max	− 0.04 ∣ 0.14	− 0.07 ∣ 0.15	− 0.14 ∣ 0.15	− 0.10 ∣ 0.19	− 0.28 ∣ 0.63	− 0.07 ∣ 0.11	− 0.09 ∣ 0.15
Median	0.0028	0.0015	0.0056	0.0049	0.0045	0.0045	0.0056
NW t-statistics (*p* Value)	4.00 (0.00)	2.90 (0.00)	3.71 (0.00)	4.23 (0.00)	1.74 (0.04)	3.98 (0.00)	3.92 (0.00)

Journal : SmallExtended 10260 | Article No : 702 | Pages : 32 | MS Code : 702 | Dispatch : 21-4-2023

Table 7 (continued)

Pre-selection metrics	$\widehat{SSD}$	$\widehat{PR}$	$\widehat{\rho}^p$	$\widehat{\rho}^r$	$\widehat{COV}^p$	$\widehat{COV}^r$	$\widehat{C}_{\Delta p}(0)$
Consistent p Value	0.00	0.00	0.00	0.00	0.05	0.00	0.00
Sharpe ratio	0.29	0.21	0.24	0.19	0.13	0.26	0.32
% positive excess returns	54.47%	51.91%	56.17%	59.15%	53.62%	57.02%	58.30%
Including commissions and cut rules (in \$)							
Mean	0.0040	0.0053	0.0117	0.0025	0.0067	0.0052	0.0102
Standard deviation	0.0272	0.0298	0.0413	0.0237	0.0680	0.0248	0.0307
Min | Max	– 0– 0.08 | 0.15	-0.10 | 0.16	-0.09 | 0.19	-0.07 | 0.12	-0.30 | 0.44	-0.07 | 0.10	-0.07 | 0.19
Median	0.0014	0.0016	0.0050	0.0019	0.0011	0.0027	0.0053
NW t-statistics (p Value)	2.34 (0.01)	2.69 (0.00)	3.14 (0.00)	1.95 (0.03)	1.48 (0.07)	2.59 (0.00)	3.58 (0.00)
Consistent p Value	0.02	0.00	0.00	0.03	0.09	0.01	0.00
Sharpe ratio	0.15	0.18	0.28	0.11	0.10	0.21	0.33
% positive excess returns	51.49%	49.79%	55.32%	54.47%	48.09%	55.32%	57.45%

The table reports the main descriptive statistics of the excess returns obtained implementing a cointegration-based pairs trading strategy where cointegration is tested by means of the two-step Engle-Granger procedure. Panel A reports the results obtained when in the first step we estimate $p_{1,t} = \mu + \beta p_{2,t} + \epsilon_t$ and compute the Spread as $Spread_t = p_{1,t} - \left(\widehat{\mu} + \widehat{\beta} p_{2,t}\right)$. Panel B reports the results obtained when the reversed regression is used instead, i.e. when the first step of the Engle-Granger procedure estimates $p_{2,t} = \widetilde{\mu} + \widetilde{\beta} p_{1,t} + \epsilon_t$ and the Spread is computed as $Spread_t = p_{2,t} - \left(\widetilde{\mu} + \widehat{\beta} p_{1,t}\right)$. Statistics are reported by different pre-selection metrics, namely: sum of squared deviations between normalized log-prices, $\widehat{SSD}$, price ratio of normalized log-prices, $\widehat{PR}$, log-prices correlation $\widehat{\rho}^p$, returns correlation $\widehat{\rho}^r$, log-prices covariance $\widehat{COV}^p$, returns covariance $\widehat{COV}^r$, and the magnitude-squared coherence between the first-difference of log prices at frequency zero $\widehat{C}_{\Delta p}(0)$. The null that the average monthly excess returns are not significantly positive is tested using Newey-West (1987) heteroskedasticity and autocorrelation robust t-statistics (associated p Values in parenthesis) and the Hansen (2005) consistent p Value, to control for the risk of data-snooping

practitioners, and not surprisingly the most widely used in the empirical literature on cointegration-based pairs trading.

Provided that $p_{1,t}$ and $p_{2,t}$, i.e. the time-series of the log-prices of the two stocks, AQ3 are $I(1)$, the first step of the Engle-Granger procedure consists in estimating the following regression:

$$p_{1,t} = \mu + \beta p_{2,t} + \epsilon_t \tag{17}$$

so as to obtain *OLS* estimates of β (and of a constant μ) and the estimated in-sample residuals $\widehat{\epsilon}_t$. In the second step of the procedure, the stationarity of $\widehat{\epsilon}_t$ is tested by means of the *ADF* test (Dickey & Fuller 1979). If $\hat{\epsilon}_t$, that represent the deviations from the long-run equilibrium, are found to be stationary, the two series are said to be cointegrated and are thus considered eligible for trading. As above, the trading is triggered whenever condition (13) is violated, where the *Spread* is computed using the estimates $\hat{\mu}$ and $\hat{\beta}$ obtained in the first step of the procedure, i.e. as:

$$Spread_t = p_{1,t} - \left(\hat{\mu} + \hat{\beta} p_{2,t}\right) \tag{18}$$

Panel A of Table 7 reports the results obtained using this procedure. The huge variability of pairs trading profitability across different pre-selection metrics is once again confirmed. The average excess returns range between 0.37% (0.27%), when pre-selection runs via the correlation between returns, up to 1% (1.31%), almost 3 (4) times larger, when the very same trading strategy is carried out on assets pre-selected using the log-prices correlation instead before (after) including commissions and cut rules. The same picture arises when the median, rather than the mean, is considered or looking at the variability of the excess returns. For instance, before commissions and cut rules are taken into account, the excess returns obtained with pairs pre-selected using returns correlation have a standard deviation (range) equal to 0.0274 (0.26), which is 3 times smaller than the one observed when pre-selection occurs based on covariance of log-prices, equal to 0.0724 (0.88).

A possible limitation of the Engle-Granger procedure is that the estimates $\hat{\mu}$ and $\hat{\beta}$, and—as a consequence—the residuals $\hat{\epsilon}_t$ of which stationarity is tested to assess cointegration, and the *Spread* used to trigger the trade, may vary based on which asset is chosen as dependent variable in the OLS regression. We thus identify a pair as eligible for trading only if the stocks are cointegrated in both directions.[16] Moreover, we repeat the analysis considering the reversed regression as the true one, i.e.:

$$p_{2,t} = \tilde{\mu} + \tilde{\beta} p_{1,t} + \varepsilon_t \tag{19}$$

and, thus, computing the *Spread* used to trigger the trades as follows:

$$Spread_t = p_{2,t} - \left(\tilde{\mu} + \tilde{\beta} p_{1,t}\right) \tag{20}$$

Results, reported in Panel B of Table 7, are largely consistent with the one reported above and once again show a huge degree of variability across the different pre-selection measures considered, thus confirming that our main conclusion is robust also to the procedure used to assess the cointegration relationship.

[16] Namely, we consider a pair cointegrated if and only if both $\hat{\epsilon}_t = p_{1,t} - \hat{\beta} p_{2,t} - \hat{\mu}$ and $\tilde{\epsilon}_t = p_{2,t} - \tilde{\beta} p_{1,t} - \tilde{\mu}$ (with $\hat{\mu}$ and $\hat{\beta}$ are the parameters estimated regressing $p_{1,t}$ on $p_{2,t}$ and $\tilde{\mu}$ and $\tilde{\beta}$ are the parameters obtained regressing $p_{2,t}$ on $p_{1,t}$) are found to be stationary. For the ADF test required, we consider the model specification featuring a constant but not a time trend and use -3.37 as critical value, assuming a level of significance equal to 5% (see MacKinnon 2010).

6 Conclusions

This study compares the profitability of a cointegration-based pairs trading strategy when pairs of US stocks are pre-selected based on seven different metrics, with the aim to reduce the computational burden entailed by cointegration tests. Although some of these metrics have already been employed in this steam of literature, to the best of our knowledge, the effect of this pre-selection on the final profitability of the pairs trading strategy has never been assessed. We also enlarge the set of pre-selection metrics considered by investigating three supplementary metrics with desirable features that, as far as we know, have never been used in this type of application.

The first take-away from our investigation is that pre-selection matters, since the profitability of the pairs trading strategy remarkably changes depending on the pre-selection metrics considered. For instance, when neither commissions nor cut rules are considered, average excess returns are not statistically significant when pre-selection is carried out by means of the covariance between log-prices or coherence, while significant in all other cases. Moreover, the average excess returns vary by a factor of 4, ranging from (a non-significant) 0.19%, generated when pairs are selected based on the coherence, up to (a significant) 0.73% when pairs are pre-selected based on the log-prices correlation. The excess returns also differ in terms of variability, whereby pre-selecting again via returns correlation as well as via SSD and PR seems to generate excess returns that are less volatile compared to those obtained when other metrics are used. Among the seven pre-selection metrics used, the ones providing the best profitability per unit of risk, as measured by the Sharpe Ratio, and the highest incidence of actually positive excess returns, are the Price Ratio and the returns correlation, respectively.

The same conclusion is achieved even after commission costs and cut rules are considered, and using a stricter definition of reversion of the Spread to the equilibrium. These differences are even more striking once the implicit costs entailed by the market impact are considered, as the pairs trading profits retain statistical significance when pre-selection runs based on log-prices correlation only, while in all the remaining cases final profits are not statistically distinguishable from zero. This is consistent with Miao (2014), who argues that pre-selection based on correlation might be beneficial to cointegration-based pairs trading in the light of the potentially complementary information it captures.

Pre-selection also impacts on the risk-profile of the observed excess returns. We find that, for all the conventional risk factors considered, results are highly disparate across the pre-selection metrics analysed. As an example, market neutrality of the pairs trading strategy is disproved if pairs are pre-selected via SSD or log-prices correlation, while supported if pre-selection is done with any other metrics considered. Similarly, the expected negative association with the momentum factor is confirmed for some metrics only. Moroever, pairs trading produce significant alphas independently of the pre-selection metric used, with the only exception of Price Ratio.

The second take-away of this analysis is that none among the three additional pre-selection measures investigated seem to produce systematically better results in terms of profitability. In other words, their features, potentially able to overcome some of the limitations of the metrics traditionally considered in the pairs trading literature, do not translate into substantial gains in terms of final profitability and, in the case of spectral coherence, also come with a higher computational burden.

Potentially interesting extensions of the analysis proposed, such as the investigation across non-US and/or non-stock markets, or a sensitivity analysis of the final profitability to arbitrarily fixed parameters (such as the opening trigger or the length of the formation period), are left for further research. AQ4

Acknowledgement The authors would like to thank Gianluca Cubadda, Stefano Grassi, Davide Pirino, and Tommaso Proietti, the participants to the Tor Vergata Ph.D Seminars and to the 30th Annual Meeting of the EFMA, as well as two anonymous referees for helpful comments and suggestions. This paper was drafted when Roberta De Luca was PhD Candidate in Economics and Finance at the University of Rome Tor Vergata. Marianna Brunetti kindly acknowledges financial support from the University of Rome Tor Vergata, Grant "Beyond Borders" (NR: E84I20000900005). The views expressed herein are those of the authors and do not involve the responsibility of the Bank of Italy. None of the authors have financial interests directly or indirectly related to this work.

Funding Open access funding provided by Università degli Studi di Roma Tor Vergata within the CRUI-CARE Agreement.

References

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

Angel J (2014) When finance meets physics: the impact of the speed of light on financial markets and their regulation. The Financial Rev 49(2):271–281

Baronyan SR, Boduroglu II, Sener E (2010) Investigation of stochastic pairs trading strategies under different volatility regimes. Manch Sch 78:114–134

Blázquez MC, De la Orden DC, Román CP (2018) Pairs trading techniques: an empirical contrast. Eur Res Manag Bus Econ 24(3):160–167

Bondarenko O (2003) Statistical arbitrage and securities prices. Rev Financial Stud 16(3):875–919

Bookstaber, R. (2007). A demon of our design. markets, hedge funds, and the perils of financial innovations. Wiley, Hoboken

Brogaard J, Hendershott T, Riordan R (2014) High-frequency trading and price discovery. Rev Financial Stud 27(8):2267–2306

Caldeira JF, Moura GV (2013) Selection of a portfolio of pairs based on cointegration: a statistical arbitrage strategy. Braz Rev Finance 11(1):49–80

Cavaliere G, De Angelis L, Rahbek A, Taylor A (2015) A comparison of sequential and information-based methods for determining the cointegration rank in heteroskedastic VAR models. Oxford Bull Econ Stat 77:106–128

Cavaliere G, De Angelis L, Rahbek A, Taylor A (2018) Determining the cointegration rank in heteroskedastic VAR models of unknown order. Economet Theor 34:349–382

Journal : SmallExtended 10260 | Article No : 702 | Pages : 32 | MS Code : 702 | Dispatch : 21-4-2023

Chen H, Chen S, Chen Z, Li F (2017) Empirical investigation of an equity pairs trading strategy. Manage Sci 65(1):370–389
Clark C (2012) How to keep markets safe in the era of high-speed trading. Chicago Fed Letter (303).
Clegg M, Krauss C (2018) Pairs trading with partial cointegration. Quant Finance 18(1):121–138
Cubadda G (1994) Testing for cointegration at any frequency using spectral methods. J. Italian Stat. Soc. 3:37–50
D'Avolio G (2002) The market for borrowing stocks. J Financ Econ 66:271–306
Dickey DA, Fuller WA (1979) Distribution of the estimators for autoregressive time series with a unit root. J Am Stat Assoc 74(366a):427–431
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.
Do B, Faff R (2010) Does simple pairs trading still work? Financ Anal J 66(4):83–95
Do B, Faff R (2012) Are pairs trading profits robust to trading costs? J Financial Res 35(2):261–287
Dunis CL, Giorgioni G, Laws J, Rudy J (2010) Statistical arbitrage and high-frequency data with an application to Eurostoxx 50 equities. Liverpool Business School, Working paper.
Ehrman DS (2006) The handbook of pairs trading: strategies using equities, options, and futures (Vol. 240). Wiley.
Emery GW, Liu QW (2002) An anlysis of the relationship between electricity and natural-gas futures prices. J Futur Mark 22(2):95–122
Engelberg J, Gao P, Jagannathan R (2009) An anatomy of pairs trading: the role of idiosyncratic news, common information and liquidity. In: Proceedings of the 3rd Singapore international conference on finance.
Engle RF, Granger CW (1987) Co-integration and error correcton: representation, estimation, and testing. Econometrica 55(2):251–276
Fama EF, French KR (2015). A five-factor asset pricing model. J Financial Econ; 116((1), 1–22.
Fama EF, French KR (1993) Common risk factors in the returns on stocks and bonds. J Financ Econ 33(1):3–56
Gatev E, Goetzmann WN, Rouwenhorst KG (2006) Pairs trading: perfomance of a relative-value arbitrage rule. Rev Financial Stud 19(3):797–827
Girma PB, Paulson AS (1999) Risk arbitrage opportunities in petroleum future spreads. J Futur Mark 19(8):931–955
Granger C, Newbold P (1974) Spurious regressions in econometrics. J Econ 2:111–120
Gregory A, Hansen B (1996) Residual-based tests for cointegration in models with regime shifts. J Econ 70:99–126
Gutierrez JA, Tse Y (2011) Illuminating the profitability of pairs trading: a test of the relative pricing efficiency of markets for water utility stocks. J Trad 6(2):50–64
Han C, He Z, Toh AJ (2022). Pairs trading via unsupervised learning. Eur J Oper Res.
Hansen PR (2005) A test for superior predictive ability. J Bus Econ Stat 23(4):365–380
Heiko J (2015) What explains the dynamics of 100 anomalies? J Bank Finance 57:65–85
Huck N (2013) The high sensitivity of pairs trading returns. Appl Econ Lett 20(14):1301–1304
Huck N, Afawubo K (2015) Pairs trading and selection methods: is cointegration superior? Appl Econ 47(6):599–613
Jacobs H, Weber M (2016) Losing sight of the trees for the forest? Atten Alloc Anom Quant Finance 16(11):1679–1693
Jegadeesh N (1990) Evidence of predictable behavior of security returns. J Finance 45:881–898
Jegadeesh N, Titman S (1993) Returns to buying winners and selling losers: implications for stock market efficiency. J Financ 48(1):65–91
Johansen S (1988) Statistical analysis of cointegration vectors. J Econ Dyn Control 12:231–254
Krauss C (2017) Statistical arbitrage pairs trading strategies; review and outlook. J Econ Surv 31(2):513–545
Kunkel R, Compton W, Beyer S (2003) The turn-of-the-month effect still lives: the international evidence. Int Rev Financ Anal 12:207–221
Lehmann B (1990) Fads, martingales and market efficiency. Quart J Econ 105:1–28
Levy D (2002) Cointegration in frequency domain. J Time Ser Anal 23:333–339
MacKinnon JG (2010) Critical values for cointegration tests. Queen's Economics Department Working Paper, No. 1227
McConnell J, Xu W (2008) Equity returns at the turn of the month. Financ Anal J 64(2):49–64

Journal : SmallExtended 10260 | Article No : 702 | Pages : 32 | MS Code : 702 | Dispatch : 21-4-2023

Miao GJ (2014) High frequency and dynamic pairs trading based on statistical arbitrage using a two stage correlation and cointegration approach. Int J Econ Financ 6(3):96–110

Newey W, West K (1987) A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. Econometrica 55(3):703–708

Papadakis G, Wysocki P (2007) Pairs trading and accounting information. Boston University and MIT Working Paper.

Phillips P (1986) Understanding spurious regressions in econometrics. J Econom 33:311–340

Rad H, Low RK, Faff RW (2016) The profitability of pairs trading strategies: distance, cointegration and copula methods. Quant Finance 16(10):1541–1558

Sharma S, Narayan P (2014) New evidence on turn-of the-month effects. J Int Finan Markets Inst Money 29(3):92–108

Simon DP (1999) The soybean crush spead; empirical evidence and trading strategies. J Futur Mark 19(3):271–289

Vidyamurthy G (2004) Pairs trading: quantitative methods and analysis. Wiley, Hoboken

Wahab M, Cohn R (1994) The gold-silver spread: integration, cointegration, predictability, and ex-ante arbitrage. J Futur Mark 14(6):709–756

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Journal:	**10260**
Article:	**702**

Author Query Form

Please ensure you fill out your response to the queries raised below and return this form along with your corrections

Dear Author

During the process of typesetting your article, the following queries have arisen. Please check your typeset proof carefully against the queries listed below and mark the necessary changes either directly on the proof/online grid or in the 'Author's response' area provided below

Query	Details Required	Author's Response
AQ1	Please confirm the Article title is correctly identified and amend if necessary.	
AQ2	Please check and confirm the organization name and street information are correct in affiliation 1 and amend if necessary.	
AQ3	Please check and confirm the layout of Table 7 is correct and amend if necessary.	
AQ4	Please check and confirm the Acknowledgement section is correct and amend if necessary.	