UNIMORE UNIVERSITÀ DEGLI STUDI DI MODENA E REGGIO EMILIA

CEFIN Centro Studi Banca e Finanza

ISSN 2282-8168

CEFIN Working Papers No 89

Pre-selection in cointegration-based pairs trading

Marianna Brunetti and Roberta De Luca

November 2022

CEFIN – Centro Studi di Banca e Finanza www.cefin.unimore.it
Dipartimento di Economia Marco Biagi – Università di Modena e Reggio Emilia
Viale Jacopo Berengario 51, 41121 MODENA (Italy) tel. 39-059.2056711
(Centralino) fax 39-059 205 6927

Pre-selection in cointegration-based pairs trading*

Marianna Brunetti

University of Rome Tor Vergata, CEFIN & CEIS

Roberta De Luca

Bank of Italy

Abstract

This paper compares the final profitability of a cointegration-based pairs trading strategy when pairs of stocks are pre-selected using seven different measures. 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, and a stricter definition of the Spread reversion to the equilibrium. Besides, the profitability of the pairs trading strategy is also found heterogeneous across the different pre-selection metrics considered in terms of exposure to the traditional risk-factors.

Keywords: pairs trading, pre-selection, cointegration, profitability

JEL Codes: G10, G12, C44, C55

* This paper was drafted when Roberta De Luca was PhD Candidate in Economics and Finance at the University of Rome Tor Vergata. The authors would like to thank Gianluca Cubadda, Tommaso Proietti and the participants to the Tor Vergata PhD Seminars for helpful comments and suggestions. 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.

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, 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 can be considered 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 & Afawubo, 2015; Rad, Low, & Faff, 2016; Blázquez, De la Orden, & Román, 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 pairs trading strategy is not clear and has never been assessed. With this paper we thus aim to fill this gap, by investigating and comparing the profitability of a cointegrated-based pairs trading strategy when pairs are pre-selected based on sever different pre-selection measures.

The rest of paper is structured as follows: Section 2 briefly summarizes the empirical contributions dealing with (distance-based and cointegration-based) pairs trading; Section 3 presents the dataset and the methodology used, illustrating the pre-selection measures considered, the trading strategies, and their final evaluation; Section 4 presents the main results, while Section 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.[2]

Gatev, Goetzmann and Rouwenhorst (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[3]; 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, Goetzmann and Rouwenhorst (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, Goetzmann and Rouwenhorst (2006) reduces if the analysis is extended to 2009, while the latter investigates the sensitivity of the profitability

[2] Being the most relevant for this work, our focus will be on distance and cointegration approaches only.
[3] The normalization is performed scaling both log-prices time series to start at 1$.

found by Gatev, Goetzmann and Rouwenhorst (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 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 ones who, to the best of our knowledge, use a different metrics are 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, Goetzmann and Rouwenhorst (2006) strategy.

As Do *et al.* (2006) highlight, that 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 intradaily) 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, Low and Faff (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, Low and Faff (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 prior sorted based on the sum of squared deviations between prices and, then, cointegration tests are performed until 20 cointegrated pairs are identified.

To be notice 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 when assets are pre-selected with different metrics. With this paper, we aim to fill this gap.

3. Data and Methodology

The empirical analysis relies on the (dividend adjusted) daily closing prices of the S&P 500 index constituents, during the period from 1st January 1998 until 30th October 2018.[4] This allows to work with extremely liquid assets, characterized by high market capitalization, and relatively low transactions costs.

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

1. Pairs pre-selection: we consider a one-year formation period during which we order pairs of stocks according to seven different measures of association or distance, described in subsection 3.1;
2. 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 subsection 3.2;
3. Pairs trading: using data from a six-month trading period, we implement the trading strategy described in subsection 3.3;
4. 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 subsection 3.4.

[4] The dataset includes all the stocks belonging to the S&P 500 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. Data were retrieved from Thomson Reuters DataStream.

3.1 Pre-selection measures

We consider seven different measures of distance - or association -, some of which have been proposed in the distance-based and cointegration-based pairs trading literature. The importance of considering more than one metric to pre-select pairs is confirmed by the results reported in subsection 3.2 below.

The first measure of association is the absolute value of the Pearson Correlation between the log-prices time series, employed in Miao (2014), that is:

$$\hat{\rho}^p = \left| \frac{\sum_{t=1}^{T}[p_{1,t} - \overline{p_1}][p_{2,t} - \overline{p_2}]}{\sqrt{\sum_{t=1}^{T}[p_{1,t} - \overline{p_1}]^2 \sum_{t=1}^{T}[p_{2,t} - \overline{p_2}]^2}} \right| \tag{1}$$

where $p_{1,t}$ and $p_{2,t}$ are the log-prices of stock 1 and 2 on day t, $\overline{p_1}$ and $\overline{p_2}$ are their corresponding sample means over the formation period, and T is the number of trading days comprising the formation period.

The second measure used to pre-select pairs of stocks is the absolute value of the Pearson Correlation Coefficient in between the log-returns, used as criterion of pairs formation by Chen, Chen, & Li (2017), that is:

$$\hat{\rho}^r = \left| \frac{\sum_{t=1}^{T}(r_{1,t} - \overline{r_1})(r_{2,t} - \overline{r_2})}{\sqrt{\sum_{t=1}^{T}(r_{1,t} - \overline{r_1})^2 \sum_{t=1}^{T}(r_{2,t} - \overline{r_2})^2}} \right| \tag{2}$$

where $r_{1,t}$ and $r_{2,t}$ are the returns on day t of stock 1 and 2, respectively obtained as difference of the stock log-prices (i.e., $r_{1,t} = p_{1,t} - p_{1,t-1}$ and $r_{2,t} = p_{2,t} - p_{2,t-1}$), $\overline{r_1}$ and $\overline{r_2}$ are their corresponding sample means over the formation period, and T is the number of trading days comprising the formation period.

Both $\hat{\rho}_{1,2}^P$ and $\hat{\rho}_{1,2}^r$, the Correlation between log-prices and returns, tend to their maximum value as the standard deviation of log-prices and of returns 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, the corresponding Covariances, between both log-prices and returns, are considered as the third and fourth pre-selection metrics, that is:

$$\widehat{COV}^p = \sum_{t=1}^{T} [p_{1,t} - \overline{p_1}][p_{2,t} - \overline{p_2}] \tag{3}$$

$$\widehat{COV}^r = \sum_{t=1}^{T} (r_{1,t} - \overline{r_1})(r_{2,t} - \overline{r_2}) \tag{4}$$

The fifth measure of association is a modified version of the Pearson Correlation Coefficient in absolute value proposed by Erdem *et al.* (2014), referred to hereafter as the "Lagged log-Prices Correlation", that is:

$$\hat{\rho}^{lag\_p} = \left| \frac{\sum_{t=1}^{T} (p_{1,t} - p_{1,t-1})(p_{2,t} - p_{2,t-1})}{\sqrt{\sum_{t=1}^{T} (p_{1,t} - p_{1,t-1})^2 \sum_{t=1}^{T} (p_{2,t} - p_{2,t-1})^2}} \right| \tag{5}$$

where, T is the number of trading days in the formation period, and $p_{1,t}$ and $p_{2,t}$ are the log-prices of stock 1 and 2 on day t. This modification exploits the first difference of log-prices time series instead of the distance with respect to their mean. According to Erdem *et al.* (2014), this should overcome the problems related to prices non-stationarity and to incorrect estimate of Pearson Correlation due to the use of average price values. As Erdem *et al.* (2014) explain, the Pearson Correlation Coefficient can fail to detect the direction of the movement in prices or returns if the variables are both above (or both below) their means, resulting in a positive relation even though the variables exhibit an opposite behavior.

The sixth measure used to pre-select pairs of stocks is the Sum of Squared Deviations between the normalized log-prices, firstly proposed as a selection criterion by Gatev *et al.* (2006) and employed for pre-selection of pairs by Rad *et al.* (2016), that is:

$$\widehat{SSD} = \sum_{t=1}^{T} (\tilde{p}_{1,t} - \tilde{p}_{2,t})^2 \tag{6}$$

where $\tilde{p}_{1,t}$ and $\tilde{p}_{2,t}$ are the normalized log-prices of stock 1 and 2 on day t, respectively, i.e. $\tilde{p}_{1,t} = \ln(P_{1,t})/\ln(P_{1,t=1})$ and $\tilde{p}_{2,t} = \ln(P_{2,t})/\ln(P_{2,t=1})$, and T is the number of trading days comprising the formation period.

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

$$\widehat{PR} = \frac{1}{T}\sum_{t=1}^{T}\frac{\tilde{p}_{1,t}}{\tilde{p}_{2,t}} \tag{7}$$

where $\tilde{p}_{1,t}$ and $\tilde{p}_{2,t}$ are the above-defined normalized log-prices of stock 1 and 2 on day t, respectively, and T is again the number of trading days in the formation period. 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. Inspired by this methodology, we employ this measure to pre-select stocks, by giving preference to those pairs with $\widehat{PR}$ closest to one.

Each metrics is computed monthly using data from the previous 1-year formation period. Stock pairs are then sorted in descending order according to measures (1) to (5), in ascending order according to SSD and by PR closest to one. 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 pairs are progressively tested for cointegration until 20 pairs of stocks whose prices are cointegrated are found. Those "top 20" pairs are thus eligible for trading during the following 6 months[5].

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). According to the methodology outlined by Vidyamurthy (2004), provided that $p_{1,t}$ and $p_{2,t}$, i.e. the time-series of the log-prices of the two stocks, are $I(1)$, the first step consists in regressing $p_{1,t}$ on $p_{2,t}$ in order to obtain OLS estimates of β (and of a constant μ) and the estimated in-sample residuals $\hat{\epsilon}$:

$$\hat{\epsilon}_t = p_{1,t} - \hat{\beta}p_{2,t} - \hat{\mu} \tag{8}$$

Then, stationarity of $\hat{\epsilon}_t$ is tested by means of the ADF test (Dickey & Fuller, 1979). If the residuals $\hat{\epsilon}_t$, 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

[5] As robustness check, we also considered the case of selecting the first 50 and 100 pairs. The results, available upon request, remain qualitatively unchanged.

trading.[6] This concludes the second step of the procedure, which – as the first one – is carried out using 1-year data in the formation period.

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 bottom panel of Table 2). With the only exception of the pairs selected using Correlation between Returns and Correlation between Lagged log-Prices, which coincide almost entirely (99%), in each same formation period the pairs selected overlap by at most 25%, and in some cases by less than 1%. This confirms the crucial impact of the pre-selection metrics used on the tradable pairs that are selected and, hence, on the final profitability of the pairs trading strategy.

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

	$\hat{\rho}^p$	$\hat{\rho}^r$	$\widehat{COV}^p$	$\widehat{COV}^r$	$\hat{\rho}^{lag\_p}$	$\widehat{SSD}$	$\widehat{PR}$
			Comparison of the cointegration tests required				
Tested pairs	27	268	259	514	266	51	551
Potential pairs	100646	100639	100646	100639	100639	100643	100642
			Comparison of the top 20 tradable pairs selected				
$\hat{\rho}^p$	1	0.25	0.10	0.12	0.26	0.10	0.02
$\hat{\rho}^r$		1	0.01	0.13	0.99	0.17	0.02
$\widehat{COV}^p$			1	0.19	0.01	0.00	0.00
$\widehat{COV}^r$				1	0.13	0.01	0.02
$\hat{\rho}^{Diff}$					1	0.17	0.02
$\widehat{SSD}$						1	0.11
$\widehat{PR}$							1

Notes: 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. $\hat{\rho}^p$ is the log-prices correlation, $\hat{\rho}^r$ is the log-returns correlation, $\widehat{COV}^p$ is the log-prices covariance, $\widehat{COV}^r$ is the log-returns covariance, $\hat{\rho}^{lag\_p}$ is the lagged log-prices correlation, $\widehat{SSD}$ is the sum of squared deviations between normalized log-prices and $\widehat{PR}$ is the price ratio of normalized log-prices.

[6] A possible limitation of the Engle and Granger procedure is the sensitivity of the results to the directionality of the regressors. We thus test for cointegration in both directions and identify a pair as eligible for trading only if the stocks are cointegrated in both directions, i.e. if both $\hat{\epsilon}_{1,t} = p_{1,t} - \hat{\beta}_1 p_{2,t} - \hat{\mu}_1$ and $\hat{\epsilon}_{2,t} = p_{2,t} - \hat{\beta}_2 p_{1,t} - \hat{\mu}_2$, (with $\hat{\mu}_1$ and $\hat{\beta}_1$are the parameters estimated regressing $p_{1,t}$ on $p_{2,t}$ and $\hat{\mu}_2$ and $\hat{\beta}_2$ are the parameters obtained regressing $p_{2,t}$ on $p_{1,t}$) are both stationary. This procedure has no impact on the identification of relatively over/under priced stocks within the pair, while it ensures that cointegration is always satisfied.

3.3 Pairs trading strategy implementation

Next, we use data from the following six-month period, the so-called trading period, and implement the self-financing pairs trading strategy whenever pricing anomalies are signaled by deviations of the out-of-sample residuals from the long-run equilibrium relationship, which is stationary and mean-reverting.

In order to do so, we define the short-term *Spread* between log-prices as:

$$Spread_t = p_{1,t} - (\hat{\mu} + \hat{\beta} p_{2,t}) \tag{9}$$

using the $\hat{\mu}$ and $\hat{\beta}$ estimates obtained in the first step of the procedure.[7] 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:

$$2\hat{\sigma} \geq Spread_t \geq -2\hat{\sigma} \tag{10}$$

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[8] (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 Figure 1.

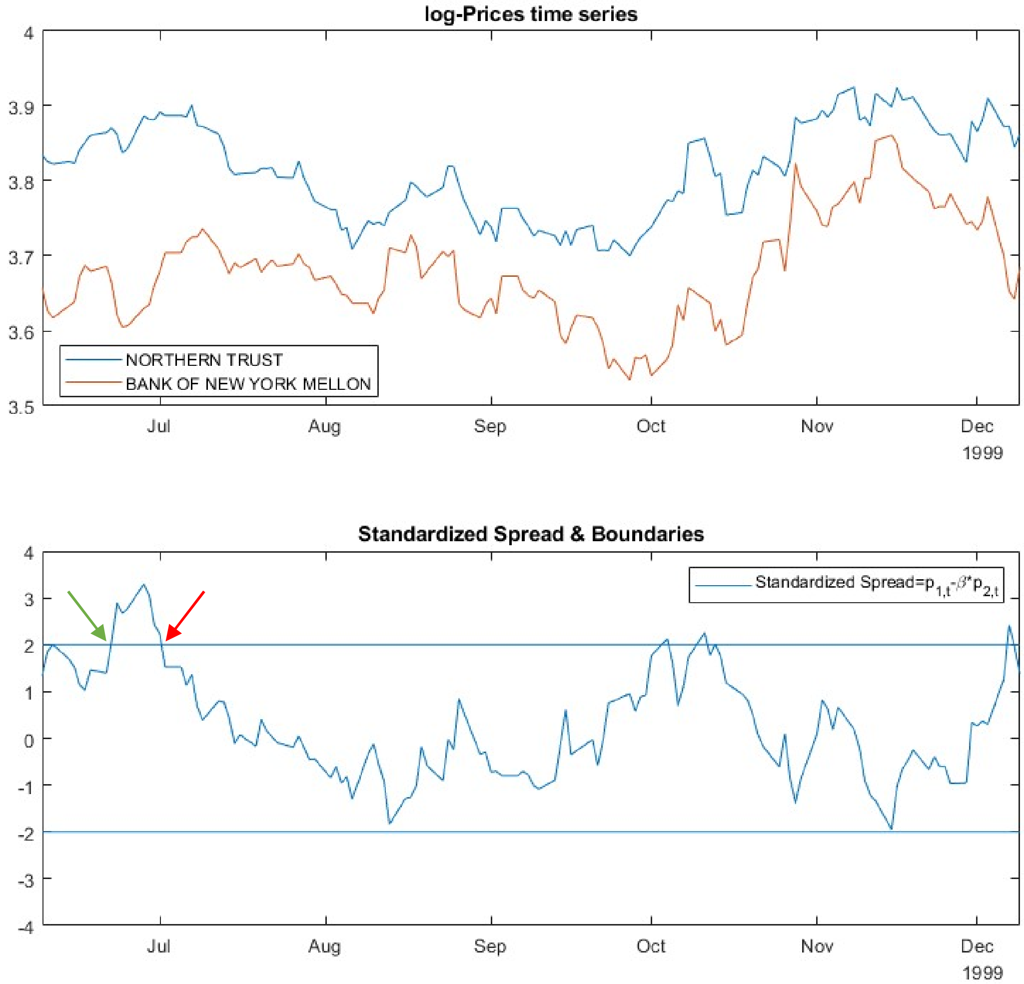
This whole procedure, from step 1 to step 3, is then repeated in rolling window fashion, by shifting the formation and trading periods one month. As a result, every month (starting

[7] The parameters used are the ones estimated regressing $p_{1,t}$ on $p_{2,t}$.

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

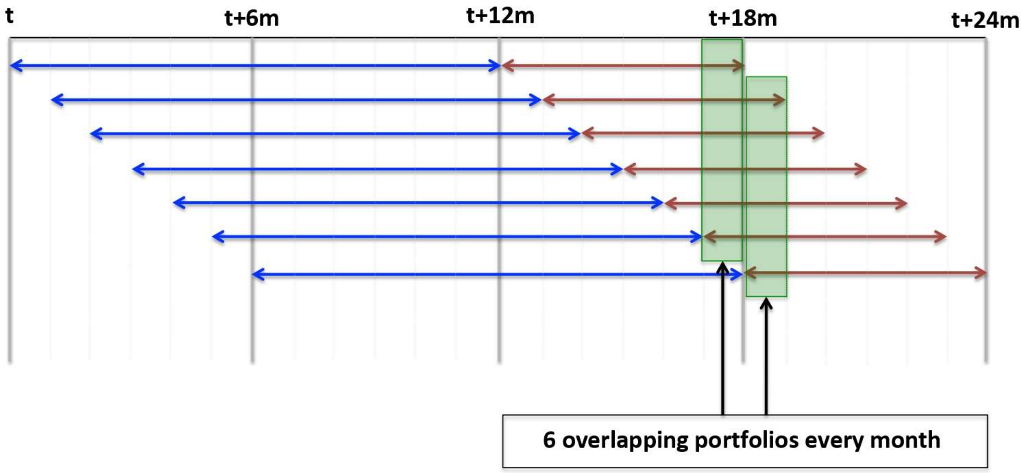
from the 6th in the sample) six overlapping portfolios of 20 pairs are generated (Figure 2). This approach mimics the payoffs that a proprietary trading desk would get delegating the management of these portfolios to six different traders whose formation and trading periods are staggered by one month (see, e.g. Gatev, Goetzmann, & Rouwenhorst, 2006 and Huck and Afawubo, 2015).

Figure 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 ±2 boundaries. The green arrow spots the opening of the trading position, while the red one spots the closing of the same trade.

Figure 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).

Our dataset comprises 235 trading months.[9] 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.

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:

$$R_{P,t} = \frac{\sum_{i=1}^{20} w_{i,t} R_{i,t}}{\sum_{i=1}^{20} w_{i,t}} \tag{11}$$

where:

- $w_{i,t}$ 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: $w_{i,t} = w_{i,t-1}(1 + R_{i,t-1}) = (1 + R_{i,1}) \dots (1 + R_{i,t-1})$
- $R_{i,t}$ is the daily mark-to-market excess return obtained from trading pair i, computed as:

[9] 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.

$$R_{i,t} = \sum_{j=1}^{2} I_{j,t}\omega_{j,t}R_{j,t} \tag{12}$$

with:

- $I_{j,t}$ 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 in day 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.

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, Goetzmann, & Rouwenhorst (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.[10]

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

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

[10] 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.

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:

$$R_{i,t} = \sum_{j=1}^{2} \left(I_{j,t}\omega_{j,t}R_{j,t} - c\omega_{j,t}\left(1 + R_{j,t}\right) \right) \tag{13}$$

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 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. [11]

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 pairs trading strategy is, in all cases, statistically significant and remarkably differs depending on the pre-selection metrics used. The average excess return obtained when pairs are pre-selected based on the Covariance between log-prices, equal to 1.63%, is almost 4 times the average net profit obtained carrying out the very same trading strategy using the Correlation between log-returns or the lagged log-prices Correlation as pre-selection metrics.

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

[11] Besides direct costs, one might also consider the implicit cost represented by the market impact. This is evaluated separately (see subsection 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 can no longer be self-financing and the final profits can no longer be interpreted as excess returns.

Covariance between log-returns (0.59%), more than doubles the lowest median excess return, obtained using the Price Ratio as pre-selection measure.

The profitability observed for the different pairs pre-selection measures is heterogeneous also in terms of variability. Pre-selecting via SSD and Covariance between log-prices provide respectively the highest and the lowest volatility of excess returns. In fact, the excess returns obtained with pairs pre-selected using SSD have a standard deviation (range) equal to 0.0207 (0.14), much smaller than the one observed when pre-selection occurs based on Covariance of log-prices, equal to 0.1086 (1.17).

Not surprisingly, profitability per unit of risk, as measured by the Sharpe Ratio, is not uniform across the different pre-selection metrics considered: the Sharpe Ratio for excess returns obtained equals 0.15 when pre-selection is based on the Covariances or on the lagged log-price Correlation, it doubles when log-prices Correlation is used to pre-select, and it further increases, reaching as much as 0.39, when pre-selection is done via SSD.

The effect of using different pre-selection measures is also evident looking at the share of trading periods (months in our case) providing a final positive excess return. In the less favorable case, i.e. when pairs are pre-selected using the Covariance between log-prices, 50.64% of the trading months eventually provide a positive excess return. This share increases to 56.60% if pre-selection uses Correlation, rather than Covariance, between log-prices and reaches its maximum when pairs are pre-selected based on SSD (62.13%).

To sum up, all our results strongly support the fact that the profitability of a pairs trading strategy remarkably differs depending on the pre-selection metrics used. Moreover, the best results in terms of profitability are obtained pre-selecting pairs via SSD, as excess returns in this case show the highest Sharpe Ratio and the highest average frequency of positive monthly excess returns.

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

Pre-selection metrics	$\hat{\rho}^p$	$\hat{\rho}^r$	$\widehat{COV}^p$	$\widehat{COV}^r$	$\hat{\rho}^{lag\_p}$	$\widehat{SSD}$	$\widehat{PR}$
Before commissions and cut rules (in $)							
Mean	0.0098	0.0041	0.0163	0.0100	0.0041	0.0080	0.0058
Standard deviation	0.0329	0.0263	0.1086	0.0687	0.0263	0.0207	0.0259
Min | Max	-0.09 | 0.14	-0.12 | 0.10	-0.31 | 0.86	-0.30 | 0.50	-0.12 | 0.10	-0.05 | 0.09	-0.09 | 0.12
Median	0.0052	0.0031	0.0040	0.0059	0.0027	0.0052	0.0026
NW t-statistics (pValue)	4.55 (0.00)	3.14 (0.00)	2.17 (0.01)	1.97 (0.02)	3.07 (0.00)	5.23 (0.00)	2.99 (0.00)
Consistent pValue	0.00	0.00	0.03	0.02	0.00	0.00	0.00
Sharpe ratio	0.30	0.16	0.15	0.15	0.15	0.39	0.22
% positive excess returns	56.60%	54.04%	50.64%	55.74%	54.04%	62.13%	52.77%
Including commissions and cut rules (in $)							
Mean	0.0093	0.0036	0.0208	0.0143	0.0035	0.0045	0.0037
Standard deviation	0.0396	0.0254	0.1141	0.0690	0.0251	0.0203	0.0255
Min | Max	-0.10 | 0.25	-0.06 | 0.13	-0.25 | 1.01	-0.14 | 0.37	-0.06 | 0.11	-0.06 | 0.08	-0.10 | 0.10
Median	0.0055	0.0024	0.0000	-0.0003	0.0022	0.0032	0.0005
NW t-statistics (pValue)	2.66 (0.00)	2.24 (0.01)	2.40 (0.01)	2.80 (0.00)	2.23 (0.01)	2.97 (0.00)	1.76 (0.04)
Consistent pValue	0.00	0.02	0.02	0.01	0.02	0.00	0.04
Sharpe ratio	0.24	0.14	0.18	0.21	0.14	0.22	0.15
% positive excess returns	55.74%	53.19%	47.66%	47.66%	52.34%	54.47%	48.51%

Notes: 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: log-prices correlation $\hat{\rho}^p$, returns correlation $\hat{\rho}^r$, log-prices covariance $\widehat{COV}^p$, returns covariance $\widehat{COV}^r$, lagged log-prices correlation $\hat{\rho}^{lag\_p}$, sum of squared deviations between normalized log-prices, $\widehat{SSD}$, and price ratio of normalized log-prices, $\widehat{PR}$. 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 pvalues in parenthesis) and the Hansen (2005) consistent pvalue, to control for the risk of data-snooping.

The impact of pre-selecting with different metrics on cointegration-based pairs trading final profitability is also revealed when commissions and cut rules are taken into account (see Panel B, Table 3). Indeed, the average monthly excess returns are found to be statistically significant and to remarkably vary across the different pre-selection metrics, ranging from 0.36% (obtained when pairs are pre-selected based on the Correlation between log-returns) to as much as almost 6 times more, 2.08%, when using the Covariance between log-prices pre-selection. Similarly, the Sharpe Ratios range between 0.14, when pre-selection is based on log-returns Correlation, and 0.24, when pairs are pre-selected via Correlation between log-prices. Finally, the frequency of favorable trading periods, i.e. of months giving a final positive excess return, also remarkably changes depending on the pre-selection metrics used: Correlation between log-prices and SSD still guarantee the highest frequencies (55.74% and 54.47%, respectively), while pre-selection based on Covariances might reduce this chance below 50%.[12]

[12] Notice that, since the effects of commissions (transaction costs) and cut rules are opposite, the overall effect on the final profitability of the strategy might be either positive or negative. Besides, notice that the rankings based e.g. on Sharpe Ratio or frequency of positive excess returns slightly change for some pre-selection metrics, meaning that the choice of pre-selection measures also matters in this direction.

All in all, the evidence reported proves that pre-selection matters, since it impacts remarkably on the profitability of the pairs trading strategy and that, among the pre-selection metrics considered, SSD and Correlation between log-prices are the ones producing the highest returns per unit of risk (mainly due to the reduced variability of the final excess returns) and the higher chances of getting final positive results.

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 results to both a stricter definition of the Spread reversion to the equilibrium and the inclusion of market impact on the final evaluation of the strategy profitability.

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 risk factors conventionally take into consideration in the pairs trading literature. 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. In this analysis, we regress the monthly excess returns (after commision and cut rules are included) on the following factors[13]:

1. Market excess return (MKT): difference between the market and 30-day Treasury bill returns;
2. Size factor (SMB): difference between small and big stock portfolios;
3. Book-to-market factor (HML): difference between value and growth stock portfolios;
4. Momentum factor (MOM): difference between last year winner and loser portfolios;

[13] 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.

5. Short-term reversal factor (STR): difference between last month winner and loser portfolios;
6. Investment factor (CMA): difference between conservative and aggressive portfolios;
7. Profitability factor (RMW): difference between robust and weak profitability portfolios.

As reported in Table 4, the size factor is the only one consistently non-significant across the pre-selection metrics considered. For all the other factors, results are highly disparate, thus proving that pre-selection impacts also on the risk-profile of the pairs trading eventual profits. The trading excess returns do correlate with the market excess returns if pairs are pre-selected via Price Ratio or Correlation between log-prices (albeit midly in this last case): hence, in these cases, the market-neutrality of the pairs trading strategy would be disproved. By constrast, if pre-selction is based on any other of the remaining metrics considered, the evidence would be in support of market neutrality. Besides, pre-selection via PR is the only metric having a strong relation with the other factors of the Fama and French (2015) five-factors model: it produces excess returns negatively correlated to the Book-to-market factor (HML) and positively correlated to investment (CMA) and profitability factors (RMW). Finally, consistently with the previous literature, the pairs trading excess returns are found to be correlated negatively with the momentum factor (MOM) and positively with the short-term reversal (ST-rev), albeit the evidence is again quite mixed in terms of statistical significance. On the one hand, the momentum parameter is strongly significant in 4 cases out of 7 (pre-selection based on Covariances, SSD or PR), midly so (at 10% level, using log-prices Correlation pre-selection) in 1 case, and not significant in the remaning cases. On the other hand, pairs trading excess returns are somewhat associated with the short-term reversal factor when pairs are pre-selected using log-returns Correlation, lagged log-prices Correlation or Price Ratio and midly so when based on SSD.

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

	$\hat{\rho}^p$	$\hat{\rho}^r$	$\widehat{COV}^p$	$\widehat{COV}^r$	$\hat{\rho}^{lag\_p}$	$\widehat{SSD}$	$\widehat{PR}$
Alpha	0.009***	0.004**	0.026***	0.013***	0.003*	0.004***	0.002
	(0.003)	(0.002)	(0.008)	(0.005)	(0.002)	(0.001)	(0.002)
Mkt	0.116*	-0.049	-0.007	0.079	-0.037	-0.009	0.098***
	(0.063)	(0.040)	(0.178)	(0.109)	(0.040)	(0.030)	(0.036)
SMB	-0.076	-0.047	-0.259	0.126	-0.044	0.056	0.095
	(0.108)	(0.070)	(0.306)	(0.188)	(0.069)	(0.052)	(0.063)
HML	-0.014	0.075	0.132	-0.308	0.067	-0.019	-0.138**
	(0.112)	(0.072)	(0.318)	(0.195)	(0.071)	(0.054)	(0.065)
MOM	-0.097*	-0.047	-0.556***	-0.271***	-0.041	-0.129***	-0.172***
	(0.058)	(0.037)	(0.163)	(0.100)	(0.037)	(0.028)	(0.033)
ST-rev	0.037	0.098**	-0.006	0.146	0.096**	0.053*	0.089**
	(0.064)	(0.041)	(0.181)	(0.111)	(0.041)	(0.031)	(0.037)
CMA	0.053	-0.079	-0.194	0.251	-0.066	0.121	0.219**
	(0.171)	(0.110)	(0.483)	(0.296)	(0.109)	(0.082)	(0.099)
RMW	-0.027	-0.072	-0.516	0.030	-0.061	0.090	0.122*
	(0.127)	(0.081)	(0.358)	(0.219)	(0.080)	(0.061)	(0.073)
Obs.	225	225	225	225	225	225	225
R^2	0.060	0.059	0.096	0.071	0.053	0.165	0.241
Adjusted R^2	0.030	0.028	0.067	0.041	0.023	0.138	0.217

Notes: 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), Momentum (MOM), Short-term Reversal (STR), Conservative minus Aggressive (CMA) and Robust minus Weak (RMW). Robust standard errors are given in parentheses. ∗significant at 10% level. ∗∗significant at 5% level. ∗∗∗significant at 1% level.

Finally, the risk factors considered are able to sweep away the significance of the intercept when pre-selection runs based on Price Ratio only. By constrast, using any other different pre-selecton metric leads to positive and statistically significant alphas, thus proving that - in such cases - the conventional risk factors are not able to fully explain the extra returns produced by the pairs trading strategy.[14]

5.2 Robustness check

We now evaluate the robustness of our results adopting an alternative definition of the reversion to the equilibrium usually employed in the literature 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, show that all the measures of profitability considered still show a degree of variation across the pre-selection measures, which to some extent is even higher than in the baseline case. For instance, the average excess returns (before the inclusion of commissions and cut rules) range from values not statistically

[14] We also estimated different specifications including the following combinations of factors: (i) three-factors model (including MKT, SMB, HML); (ii) three-factors model augmented with MOM and ST-rev; and (iii) five-factors model (including MKT, SMB, HML, CMA, RMW). Results, available upon request, remain qualitatively unchanged.

distinguishable from 0 to 0.43%, when pre-selecting with SSD, and as much as 0.84%, when pre-selection is based on Covariance between log-prices. Similarly, the Sharpe Ratios range between 0.00001, when using Covariance between log-returns pre-selection, and 0.27, when pairs are pre-selected based on SSD. Finally, the frequency of favorable trading periods, i.e. of months providing a final positive excess return, is highest when pre-selection is done using the Correlation between log-prices and SSD and lowest (and lower than 50%) when pairs are pre-selected based on Covariances between log-prices.

Table 5 – Pairs trading robustness check monthly excess returns, by pre-selection metrics.

Pre-selection metrics	$\hat{\rho}^p$	$\hat{\rho}^r$	$\widehat{COV}^p$	$\widehat{COV}^r$	$\hat{\rho}^{lag\_p}$	$\widehat{SSD}$	$\widehat{PR}$
			Before commissions and cut rules (in $)				
Mean	0.0069	0.0003	0.0084	0.00001	0.0002	0.0043	0.0012
Standard deviation	0.0400	0.0211	0.1047	0.0594	0.0211	0.0161	0.0206
Min | Max	-0.08 | 0.46	-0.06 | 0.10	-0.33 | 0.77	-0.33 | 0.31	-0.07 | 0.10	-0.05 | 0.07	-0.08 | 0.07
Median	0.0035	0.0006	-0.0015	0.0009	0.0004	0.0024	0.0007
NW t-statistics (pValue)	2.82 (0.00)	0.27 (0.40)	1.47 (0.07)	0.00 (0.50)	0.16 (0.44)	3.79 (0.00)	0.89 (0.19)
Consistent pValue	0.00	0.41	0.10	1.00	0.44	0.00	0.21
Sharpe ratio	0.17	0.01	0.08	0.00001	0.01	0.27	0.06
% positive excess returns	54.04%	49.79%	46.81%	48.09%	49.36%	54.04%	49.36%
			Including commissions and cut rules (in $)				
Mean	0.0113	0.0055	0.0156	0.0178	0.0052	0.0031	0.0021
Standard deviation	0.0651	0.0413	0.1204	0.1711	0.0412	0.0183	0.0245
Min | Max	-0.14 |	-0.06 |	-0.26 |	-0.20 |	-0.05 |	-0.07 |	-0.10 |
Median	0.0014	0.0036	-0.0008	0.0046	0.0028	0.0020	0.0022
NW t-statistics (pValue)	2.63 (0.00)	1.86 (0.03)	1.99 (0.02)	1.47 (0.07)	1.81 (0.04)	2.07 (0.02)	1.44 (0.08)
Consistent pValue	0.02	0.04	0.03	0.09	0.06	0.01	0.08
Sharpe ratio	0.17	0.13	0.13	0.10	0.13	0.17	0.09
% positive excess returns	50.64%	55.74%	47.66%	52.77%	55.32%	53.19%	51.06%

Notes: 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: log-prices correlation $\hat{\rho}^p$, returns correlation $\hat{\rho}^r$, log-prices covariance $\widehat{COV}^p$, returns covariance $\widehat{COV}^r$, lagged log-prices correlation $\hat{\rho}^{lag\_p}$, sum of squared deviations between normalized log-prices, $\widehat{SSD}$, and price ratio of normalized log-prices, $\widehat{PR}$. 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 pvalues in parenthesis) and the Hansen (2005) consistent pvalue, to control for the risk of data-snooping.

In general, imposing this stricter condition to unwound the positions on the assets implies a longer duration of each trade, coupled with a reduction in the number of actual trades, thus sensibly reducing the final impact of commissions. Indeed, the average excess returns after commissions and cut rules are in most cases (exception are pre-selection based on the Covariance between log prices, SSD and PR) higher than in the baseline case. This further confirms the dissimilar impact on the profitability of pairs trading of using different pre-section metrics.

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 then 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 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 a remarkably reduction, of about \$0.006. This effect is again not uniform across the pre-selection metrics: for instance, net profits are reduced by about \$0.0081 when pre-selection is based on the log-prices Covariance and of about \$0.0046 when pairs are pre-selected using SSD. Moreover, 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 or log-returns) Covariance only. In all the remaining cases final profits are not statistically distinguishable from 0, further confirming the relevant impact on the eventual pairs trading profitability of using different metrics for pairs pre-selection.

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

Pre-selection metrics	$\hat{\rho}^p$	$\hat{\rho}^r$	$\widehat{COV}^p$	$\widehat{COV}^r$	$\hat{\rho}^{lag\_p}$	$\widehat{SSD}$	$\widehat{PR}$
Average net profit before market impact	0.0093	0.0036	0.0208	0.0143	0.0035	0.0045	0.0037
Average net profit after market impact	0.0037	-0.0014	0.0127	0.0073	-0.0015	-0.0001	-0.0012
Standard deviation	0.0386	0.0245	0.1115	0.0672	0.0242	0.0196	0.0246
Min | Max	-0.11 | 0.24	-0.06 | 0.11	-0.26 | 0.97	-0.15 | 0.35	-0.06 | 0.09	-0.07 | 0.07	-0.10 | 0.09
NW t-statistics (pValue)	1.10 (0.14)	-0.93 (0.82)	1.54 (0.06)	1.53 (0.06)	-1.06 (0.86)	-0.09 (0.53)	-0.61 (0.73)
Consistent pValue	0.11	1.00	0.08	0.07	1.00	1.00	1.00
% positive profits	0.49	0.42	0.46	0.44	0.43	0.43	0.42

Notes: 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: log-prices correlation $\hat{\rho}^p$, returns correlation $\hat{\rho}^r$, log-prices covariance $\widehat{COV}^p$, returns covariance $\widehat{COV}^r$, lagged log-prices correlation $\hat{\rho}^{lag\_p}$, sum of squared deviations between normalized log-prices, $\widehat{SSD}$, and price ratio of normalized log-prices, $\widehat{PR}$. 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 pvalues in parenthesis) and the Hansen (2005) consistent pvalue, to control for the risk of data-snooping.

6. Conclusions

This study compares the profitability of a cointegration-based pairs trading strategy when pairs of assets are pre-selected based on seven different measures of distance or association, which aimed 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.

Our results show 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, the average excess returns may vary by a factor of almost 4, ranging from 0.41%, generated when pairs are selected based on the Correlation between log-returns, to 1.63% when pairs are pre-selected based on the Covariance on (log-) prices. The excess returns also differ in terms of variability, whereby pre-selecting via SSD seems to generate excess returns that are much less volatile compared to those obtained when pre-selection occurs based on (log-) prices Covariance. 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 SSD and Correlation between log-prices. These results are retrieved even after commission costs and cut rules are considered, and are robust to 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 for two pre-selection metrics only, while in all the remaining cases final profits are not statistically distinguishable from zero.

Pre-selection seems to impact also on the risk-profile of the observed excess returns; in fact, for all the conventional risk factors considered but size, 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 Price Ratio or log-prices Correlation, while supported if pre-selection runs with any other metrics considered. Moroever, pairs trading produce significant alphas independently of the pre-selection metric used, with the only exception of pre-selection based on Price Ratio.

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.

References

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

Avellaneda, M., & Lee, J. H. (2010). Statistical arbitrage in the US equities market. *Quantitative Finance, 10*(7), 761-782.

Baronyan, S. R., Boduroglu, I. I., & Sener, E. (2010). Investigation of stochastic pairs trading strategies under different volatility regimes. *The Manchester School, 78*, 114-134.

Blázquez, M. C., De la Orden, D. C., & Román, C. P. (2018). Pairs trading techniques: An empirical contrast. *European Research on Management and Business Economics, 24*(3), 160-167.

Bogomolov, T. (2011). Pairs trading in the land down under. *In Finance and Corporate Governance Conference.*

Bondarenko, O. (2003). Statistical arbitrage and securities prices. *The Review of Financial Studies, 16*(3), 875-919.

Bookstaber, R. (2007). *A Demon of Our Design. Markets, Hedge Funds, and the Perils of Financial Innovations.* John Wiley and Sons.

Caldeira, J. F., & Moura, G. V. (2013). Selection of a portfolio of pairs based on cointegration: a statistical arbitrage strategy. *Brazilian Review of Finance, 11*(1), 49-80.

Chen, H., Chen, S., Chen, Z., & Li, F. (2017). Empirical investigation of an equity pairs trading strategy. *Management Science, 65*(1), 370-389.

Dickey, D. A., & Fuller, W. A. (1979). Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American statistical association, 74*(366a), 427-431.

Do, B., & Faff, R. (2010). Does simple pairs trading still work? *Financial Analysts Journal, 66*(4), 83-95.

Do, B., & Faff, R. (2012). Are pairs trading profits robust to trading costs? *Journal of Financial Research, 35*(2), 261-287.

Do, B., Faff, R., & Hamza, K. (2006). A new approach to modeling and estimation for pairs trading. *In Proceedings of 2006 financial management association European conference*, *Vol. 1*, p. 87-99.

Dunis, C. L., 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, D. S. (2006). *The handbook of pairs trading: strategies using equities, options, and futures* (Vol. 240). John Wiley & Sons.

Emery, G. W., & Liu, Q. W. (2002). An anlysis of the relationship between electricity and natural-gas futures prices. *Journal of Futures Markets, 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. *Proceedings of the 3rd Singapore International Conference on Finance.*

Engle, R. F., & Granger, C. W. (1987). Co-integration and error correcton: Representation, estimation, and testing. *Econometrica, 55*(2), 251-276.

Erdem, O., Ceyhan, E., & Varli, Y. (2014). A new correlation coefficient for bivariate time-series data. *Physica A: Statistical Mechanics and its Applications, 414*, 274-284.

Fama, E. F., & French, K. R. (1993). Common risk factors in the returns on stocks and bonds. *Journal of financial economics, 33*(1), 3-56.

Fama, E. F., & French, K. R. (2015). A five-factor asset pricing model. *Journal of financial economics, 116((*1), 1-22.

Galenko, A., Popova, E., & Popova, I. (2012). Trading in the presence of cointegration. *The Journal of Alternative Investments, 15*(1), 85-97.

Gatev, E., Goetzmann, W. N., & Rouwenhorst, K. G. (2006). Pairs trading: perfomance of a relative-value arbitrage rule. *Review of Financial Studies, 19*(3), 797-827.

Girma, P. B., & Paulson, A. S. (1999). Risk arbitrage opportunities in petroleum future spreads. *Journal of Futures Markets, 19*(8), 931-955.

Gutierrez, J. A., & Tse, Y. (2011). Illuminating the profitability of pairs trading: a test of the relative pricing efficiency of markets for water utility stocks. *The Journal of Trading, 6*(2), 50-64.

Hansen, P. R. (2005). A test for superior predictive ability. *Journal of Business & Economic Statistics, 23*(4), 365-380.

Huck, N. (2013). The high sensitivity of pairs trading returns. *Applied Economics Letters, 20*(14), 1301-1304.

Huck, N., & Afawubo, K. (2015). Pairs trading and selection methods: is cointegration superior? *Applied Economics, 47*(6), 599-613.

Jacobs, H., & Weber, M. (2016). Losing sight of the trees for the forest? Attention allocation and anomalies. *Quantitative Finance, 16*(11), 1679-1693.

Jegadeesh, N., & Titman, S. (1993). Returns to buying winners and selling losers: Implications for stock market efficiency. *The Journal of finance, 48*(1), 65-91.

Jegadeesh, N., & Titman, S. (2001). Profitability of momentum strategies: An evaluation of alternative explanations. *The Journal of finance, 56*(2), 699-720.

Johansen, S. (1988). Statistical Analysis of Cointegration Vectors. *Journal of Economic Dynamics & Control, 12*, 231-254.

Krauss, C. (2017). Statistical arbitrage pairs trading strategies; review and outlook. *Journal of Economic Surveys, 31*(2), 513-545.

Lin, Y. X., McCrae, M., & Gulati, C. (2006). Loss protection in pairs trading through minimum profit bounds: A cointegration approach. *Advances in Decision Sciences.*

Miao, G. J. (2014). High frequency and dynamic pairs trading based on statistical arbitrage using a two stage correlation and cointegration approach. *International Journal of Economics and Finance, 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.*

Puspaningrum, H., Lin, Y. X., & Gulati, C. M. (2010). Finding the optimal pre-set boundaries for pairs trading strategy based on cointegration technique. *Journal of Statistical Theory and Practice, 4*(3), 391-419.

Rad, H., Low, R. K., & Faff, R. W. (2016). The profitability of pairs trading strategies: distance, cointegration and copula methods. *Quantitative Finance, 16*(10), 1541-1558.

Sharma, A., & Seth, N. (2012). Literature review of stock market integration: a global perspective. *Qualitative Research in Financial Markets, 4*(1), 84-122.

Shiller, R. J. (1989). Comovements in stock prices and comovements in dividends. *The Journal of Finance, 44*(3), 719-729.

Simon, D. P. (1999). The soybean crush spead; empirical evidence and trading strategies. *Journal of Futures Markets, 19*(3), 271-289.

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

Wahab, M., & Cohn, R. (1994). The gold-silver spread: integration, cointegration, predictability, and ex-ante arbitrage. *Journal of Futures Markets, 14*(6), 709-756.

UNIMORE UNIVERSITÀ DEGLI STUDI DI MODENA E REGGIO EMILIA CEFIN Centro Studi Banca e Finanza

ISSN 2282-8168

88 *ESG compliant optimal portfolios: The impact of ESG constraints on portfolio optimization in a sample of European stocks?* by Bertelli, B. and Torricelli, C. (October 2022).

87 *ESG screening strategies and portfolio performance: how do they fare in periods of financial distress?,* by Bertelli, B. and Torricelli, C. (June 2022).

86 *Climate Stress Test: bad (or good) news for the market? An Event Study Analysis on Euro Zone Banks,* by Torricelli, C. and Ferrari, F. (May 2022).

85 *Social Bonds and the "Social Premium"*, by Torricelli, C. and Pellati, E. (April 2022).

84 *Do ESG Investments Mitigate ESG Controversies? Evidence From International Data,* by Brighi, P., Della Bina, A.C.F. and Venturelli, V. (March 2022).

83 *The market price of greenness. A factor pricing approach for Green Bonds,* by Bertelli, B., Boero, G. and Torricelli, C. (June 2021*).*

82 *Le offerte pubbliche di acquisto a venti anni dal Testo Unico della Finanza (Tender Offers in Italy Twenty Years After the Unified Finance Law),* by Iocca, M.G. and Ferretti, R. (January 2021*).*

81 *Financial fragility across Europe and the US: The role of portfolio choices, household features and economic-institutional setup,* by Brunetti, M., Giarda, E. and Torricelli, C. (April 2020).

80 *Birds of a feather flock together and get money from the crowd,* by Venturelli, V., Gallo, G. and Pedrazzoli, A. (December 2019).

79 *The effect of the Fed zero-lower bound announcement on bank profitability and diversification,* by Landi, A., Sclip, A. and Venturelli, V. (December 2019).

78 *Portfolio choice: Evidence from new-borns,* by Arnaboldi, F. and Gioia, F. (October 2019).

77 *Behind the success of dominated personal pension plans: sales force and financial literacy factors,* by Marotta, G. (October 2019).

76 *Short-term household income mobility before and after the Great Recession: A four country study,* by Casabianca, E.J. and Giarda, E. (June 2019).

75 *The impact of Fundamental Review of Trading Book: A preliminary assessment on a stylized portfolio,* by Pederzoli, C. and Torricelli, C. (February 2019).

74 *Security-voting structure and equity financing in the Banking Sector: 'One Head-One Vote' versus 'One Share-One Vote',* by Ferretti, R., Pattitoni, P. and Castelli, A. (January 2019).

73 *The Financial Decisions of Immigrant and Native Households: Evidence from Italy,* by Bertocchi, G., Brunetti, M. and Zaiceva, A. (January 2019).

72 *Why choosing dominated personal pension plans: sales force and financial literacy effects*, by Marotta, G. (July 2018).

71 *Spending Policies of Italian Banking Foundations,* by Pattarin, F. (July, 2018).

70 *The Forecasting Performance of Dynamic Factors Models with Vintage Data,* by Di Bonaventura, L., Forni, M. and Pattarin, F. (July 2018).

69 *Housing Market Shocks in Italy: a GVAR Approach,* by Cipollini, A. and Parla, F. (April, 2018).

68 *Customer Complaining and Probability of Default in Consumer Credit*, by Cosma, S., Pancotto, F. and Vezzani, P. (March, 2018).

67 *Is Equity Crowdfunding a Good Tool for Social Enterprises?, by Cosma, S., Grasso, A.G., Pagliacci, F. and Pedrazzoli, A. (February, 2018).*

66 *Household Preferences for Socially Responsible Investments, by Rossi. M.C., Sansone, D., Torricelli, C. and van Soest, A. (February 2018).*

65 *Market-Book Ratios of European Banks: What does Explain the Structural Fall?, by Ferretti, R.,Gallo, G., Landi, A. and Venturelli. V. (January 2018).*

CEFIN – Centro Studi di Banca e Finanza www.cefin.unimore.it
Dipartimento di Economia Marco Biagi – Università di Modena e Reggio Emilia
Viale Jacopo Berengario 51, 41121 MODENA (Italy) tel. 39-059.2056711
(Centralino) fax 39-059 205 6927

UNIMORE CEFIN
UNIVERSITÀ DEGLI STUDI DI MODENA E REGGIO EMILIA
Centro Studi Banca e Finanza

ISSN 2282-8168

64 *Past Income Scarcity and Current Perception of Financial Fragility, by Baldini, M., Gallo, G. and Torricelli, C. (December 2017).*

63 *How does Financial Market Evaluate Business Models? Evidence from European Banks, by Cosma, S., Ferretti, R.,Gualandri, E., Landi, A. and Venturelli, V. (May 2017)*

62 *Individual Heterogeneity and Pension Choices. How to Communicate an Effective Message?, by Gallo, G., Torricelli, C. and van Soest, A. (April 2017)*

61 *The Risk-Asymmetry Index,* by Elyasiani, E., Gambarelli L. and Muzzioli, S.(December 2016)

60 *Strumenti per il sostegno finanziario di famiglie e microimprese: il caso italiano (Policies to help financially vulnerable Italian households and micro-businesses), by Cotterli, S. (December 2016)*

59 *A test of the Behavioral versus the Rational model of Persuasion in Financial Advertising, by Ferretti, R., Pancotto, F. and Rubaltelli, E. (May 2016)*

58 *Financial connectedness among European volatility risk premia, by Cipollini, A., Lo Cascio, I. and Muzzioli, S. (December 2015)*

57 *The effectiveness of insider trading regulations. The case of the Italian tender offers, by Ferretti, R., Pattitoni, P. and Salinas, A. (September 2015)*

56 *Introducing Aggregate Return on Investment as a Solution for the Contradiction between some PME Metrics and IRR, by Altshuler D. and Magni, C.A. (September 2015)*

55 *'It's a trap!' The degree of poverty persistence in Italy and Europe, by Giarda, E. and Moroni, G. (September 2015)*

54 *Systemic risk measures and macroprudential stress tests. An assessment over the 2014 EBA exercise. by Pederzoli, C. and Torricelli,C. (July 2015)*

53 *Emotional intelligence and risk taking in investment decision-making, by Rubaltelli, E., Agnoli, S., Rancan, M.and Pozzoli, T. (July 2015)*

52 *Second homes: households' life dream or wrong investment?, by Brunetti, M. and Torricelli, C. (May 2015)*

51 *Pseudo-naïve approaches to investment performance measurement, by Magni, C.A. (February, 2015)*

50 *Monitoring systemic risk. A survey of the available macroprudential Toolkit, by Gualandri, E. and Noera, M. (November 2014).*

49 *Towards a macroprudential policy in the EU: Main issues, by Gualandri, E. and Noera, M. (November 2014).*

48 *Does homeownership partly explain low participation in supplementary pension schemes?, by Santantonio, M., Torricelli, C., and Urzì Brancati M.C., (September 2014)*

47 *An average-based accounting approach to capital asset investments: The case of project finance, by Magni, C.A. (September 2014)*

46 *Should football coaches wear a suit? The impact of skill and management structure on Serie A Clubs' performance,* by Torricelli, C., Urzì Brancati M.C., and Mirtoleni, L. (July 2014)

45 *Family ties: occupational responses to cope with a household income shock,* by Baldini, M., Torricelli, C., Urzì Brancati M.C. (April 2014)

44 *Volatility co-movements: a time scale decomposition analysis,* by Cipollini, I., Lo Cascio I., Muzzioli. S. (November 2013)

UNIMORE CEFIN
UNIVERSITÀ DEGLI STUDI DI MODENA E REGGIO EMILIA
Centro Studi Banca e Finanza

ISSN 2282-8168

43 *The effect of revenue and geographic diversification on bank performance,* by Brighi, P., Venturelli, V. (October 2013)
42 *The sovereign debt crisis: the impact on the intermediation model of Italian banks,* by Cosma,S., Gualandri, E. (October 2013)
41 *The financing of Italian firms and the credit crunch: findings and exit strategies,* by Gualandri, E., Venturelli, V. (October 2013)
40 *Efficiency and unbiasedness of corn futures markets: New evidence across the financial crisis,* by Pederzoli, C., Torricelli, C. (October 2013)
39 *La regolamentazione dello short selling: effetti sul mercato azionario italiano (Short selling ban: effects on the Italian stock market)*, by Mattioli L., Ferretti R. (August 2013)
38 *A liquidity risk index as a regulatory tool for systematically important banks? An empirical assessment across two financial crises*, by Gianfelice G., Marotta G., Torricelli C. (July 2013)
37 *Per un accesso sostenibile delle Pmi al credito (A sustainable access to credit for SMEs),* by Marotta, G. (May 2013)
36 *The unavoidable persistence of forum shopping in the Insolvency Regulation, by Mucciarelli, F.M. (April 2013)*
35 *Rating Triggers, Market Risk and the Need for More Regulation, by Parmeggiani, F. (December 2012)*
34 *Collateral Requirements of SMEs: The Evidence from Less–Developed Countries, by Hanedar, E.H. Broccardo, E. and Bazzana, F. (November 2012)*
33 *Is it money or brains? The determinants of intra-family decision power, by Bertocchi, G., Brunetti, M. and Torricelli, C. (June 2012)*
32 *Is financial fragility a matter of illiquidity? An appraisal for Italian households, by Brunetti, M, Giarda, E. and Torricelli, C. (June 2012)*
31 *Attitudes, personality factors and household debt decisions: A study of consumer credit, by Stefano Cosma and Francesco Pattarin (February 2012)*
30 *Corridor implied volatility and the variance risk premium in the Italian market, by Silvia Muzzioli (November 2011)*
29 *Internal Corporate Governance and the Financial Crisis: Lessons for Banks, Regulators and Supervisors, by Gualandri, E., Stanziale, A. and Mangone, E. (November 2011)*
28 *Are defined contribution pension schemes socially sustainable? A conceptual map from a macroprudential perspective, by Marotta, G. (October 2011)*
27 *Basel 3, Pillar 2: the role of banks' internal governance and control function, by Gualandri, E. (September 2011)*
26 *Underpricing, wealth loss for pre-existing shareholders and the cost of going public: the role of private equity backing in Italian IPOs*, by Ferretti, R. and Meles, A. (April 2011)
25 *Modelling credit risk for innovative firms: the role of innovation measures*, by Pederzoli C., Thoma G., Torricelli C. (March 2011)
24 *Market Reaction to Second-Hand News: Attention Grabbing or Information Dissemination?*, by Cervellati E.M., Ferretti R., Pattitoni P. (January 2011)
23 *Towards a volatility index for the Italian stock market*, by Muzzioli S. (September 2010)

UNIMORE CEFIN
UNIVERSITÀ DEGLI STUDI DI MODENA E REGGIO EMILIA Centro Studi Banca e Finanza

ISSN 2282-8168

22 *A parsimonious default prediction model for Italian SMEs*, by Pederzoli C., Torricelli C. (June 2010)

21 *Average Internal Rate of Return and investment decisions: a new perspective,* by Magni C.A. (February 2010)

20 *The skew pattern of implied volatility in the DAX index options market,* by Muzzioli S. (December 2009)

19 *Accounting and economic measures: An integrated theory of capital budgeting,* by Magni C.A. (December 2009)

18 *Exclusions of US-holders in cross-border takeover bids and the principle of equality in tender offers*, by Mucciarelli F. (May 2009).

17 *Models for household portfolios and life-cycle allocations in the presence of labour income and longevity risk*, by Torricelli C. (March 2009)

16 *Differential evolution of combinatorial search for constrained index tracking*, by Paterlini S, Krink T, Mittnik S. (March 2009)

15 *Optimization heuristics for determining internal rating grading scales*, by Paterlini S, Lyraa M, Pahaa J, Winker P. (March 2009)

14 *The impact of bank concentration on financial distress: the case of the European banking system*, by Fiordelisi F, Cipollini A. (February 2009)

13 *Financial crisis and new dimensions of liquidity risk: rethinking prudential regulation and supervision*, by Landi A, Gualandri E, Venturelli V. (January 2009)

12 *Lending interest rate pass-through in the euro area: a data-driven tale*, by Marotta G. (October 2008)

11 *Option based forecast of volatility: an empirical study in the Dax index options market*, Muzzioli S. (May 2008)

10 *Lending interest rate pass-through in the euro area, by Marotta G. (March 2008)*

9 *Indebtedness, macroeconomic conditions and banks' losses: evidence from Italy, by Torricelli C, Castellani S, Pederzoli C. (January 2008)*

8 *Is public information really public? The role of newspapers*, Ferretti R, Pattarin F. (January 2008)

7 *Differential evolution of multi-objective portfolio optimization*, by Paterlini S, Krink T. (January 2008)

6 *Assessing and measuring the equity gap and the equity*, by Gualandri E, Venturelli V. (January 2008)

5 *Model risk e tecniche per il controllo dei market parameter*, Torricelli C, Bonollo M, Morandi D, Pederzoli C. (October 2007)

4 *The relations between implied and realised volatility, are call options more informative than put options? Evidence from the Dax index options market*, by Muzzioli S. (October 2007)

3 *The maximum LG-likelihood method: an application to extreme quantile estimation in finance*, by Ferrari D., Paterlini S. (June 2007)

2 *Default risk: Poisson mixture and the business cycle*, by Pederzoli C. (June 2007)

1 *Population ageing, household portfolios and financial asset returns: a survey of the literature*, by Brunetti M. (May 2007)

UNIMORE CEFIN

UNIVERSITÀ DEGLI STUDI DI MODENA E REGGIO EMILIA

Centro Studi Banca e Finanza

CEFIN – Centro Studi di Banca e Finanza www.cefin.unimore.it
Dipartimento di Economia Marco Biagi – Università di Modena e Reggio Emilia
Viale Jacopo Berengario 51, 41121 MODENA (Italy) tel. 39-059.2056711
(Centralino) fax 39-059 205 6927