Selection of a Portfolio of Pairs Based on Cointegration: A Statistical Arbitrage Strategy

(Seleção de uma Carteira de Pares de Ações Usando Cointegração: Uma Estratégia de Arbitragem Estatística)

João Frois Caldeira*
Gulherme Valle Moura**

Abstract

Statistical arbitrage strategies, such as pairs trading and its generalizations, rely on the construction of mean-reverting spreads with a certain degree of predictability. This paper applies cointegration tests to identify stocks to be used in pairs trading strategies. In addition to estimating long-term equilibrium and to model the resulting residuals, we select stock pairs to compose a pairs trading portfolio based on an indicator of profitability evaluated in-sample. The profitability of the strategy is assessed with data from the São Paulo stock exchange ranging from January 2005 to October 2012. Empirical analysis shows that the proposed strategy exhibit excess returns of 16.38% per year, Sharpe Ratio of 1.34 and low correlation with the market.

Keywords: statistical arbitrage; pairs trading; cointegration; market neutral strategy.

JEL codes: C53; E43; G17.

Resumo

Estratégias de arbitragem estatística como *pairs trading* e suas generalizações dependem da construção de *spreads* estacionários com certo grau de previsibilidade. Este artigo aplica testes de cointegração para identificar ativos para serem usados em estratégias de *pairs trading*. Além de estimar o equilíbrio de longo prazo e

de modelar os resíduos resultantes, pares de ações são selecionados baseados em um indicador de lucratividade para compor um portfólio de pares. O retorno da estratégia é avaliado com dados diários da Bovespa durante o período de janeiro de 2005 até outubro de 2012. A análise empírica mostra que a estratégia proposta obtém excessos de retorno da ordem de 16.38% ao ano, índice de Sharpe de 1.34 e baixa correlação com o Ibovespa.

Palavras-chave: arbitragem; pairs trading; cointegração; estratégias neutras.

1. Introduction

The motivation for statistical arbitrage techniques has its roots in works that preach predictability of stock prices and existence of long term relations in the stock markets. In recent years, the notion of mean reversion has received a considerable amount of attention in the financial literature. Since future observations of a mean-reverting time series can potentially be forecasted using historical data, this literature challenges the stylized fact in financial economics which says that the stock prices shall be decribed by independent random walk processes; what would automatically imply no predictability in the stock prices (see, for example Lo & MacKinlay, 1988, 1997, Guidolin *et al.*, 2009). A number of studies have also examined the implications of mean reversion on portfolio allocation and asset management; see Barberis (2000), Carcano *et al.* (2005), Serban (2010) and Triantafyllopoulos & Montana (2011) for recent works. Active asset allocation strategies based on mean-reverting portfolios, which generally fall under the umbrella of statistical arbitrage, have been used by investment banks and hedge funds for several years. Possibly the simplest of such strategies consists of a portfolio of only two assets, as in pairs trading. This trading approach consists in going long on a certain asset while shorting another asset in such a way that the resulting portfolio has no net exposure to broad market moves. In this sense, the strategy is often described as market neutral. For further discussions on statistical arbitrage approaches based on mean-reverting spreads and many illustrative numerical examples the reader is referred to Pole (2007) and Vidyamurthy (2004).

Pairs trading is a statistical arbitrage strategy designed to exploit short-term deviations from a long-run equilibrium between two stocks. Traditional methods of pairs trading have sought to identify trading pairs based on correlation and other non-parametric decision rules. This study selects trading pairs based on the presence of a cointegrating relationship between two stocks. Cointegration enables us to combine the two stocks in a certain

linear combination so that the combined portfolio is a stationary process. If two stocks share a long-run equilibrium relationship, then deviations from this equilibrium are only short-term and are expected to die out in future periods. To profit from this relative misspricing, the trade is opened by buying the stock which is bellow the long-run equilibrium, and selling (short) the equity which is above it. The pair trade is then closed by reversing the opening transactions once the pair reverts to its expected value. The long-short transactions are constructed to yield a net position of zero.

In order to reduce risk in pairs-trading strategies, it is interesting to open many trades all with a very short holding time, hoping to diversify the risk of each trade. According to Avellaneda & Lee (2010), the pairs trading strategy is the "ancestor" of statistical arbitrage. The term "statistical arbitrage" encompasses a variety of investment strategies whose principal characteristic is the use of statistical tools to generate excess returns. Desired characteristics of this class of strategies is market neutrality (low market correlations), and signal generation based on rules rather than fundamentals.

It is well known that pairs trading is a common strategy among many hedge funds. However, there is not a significant amount of academic literature devoted to it due to its proprietary nature. For a review of some of the existing academic models, see Poterba & Summers (1988), Lo & MacKinlay (1990), Gatev *et al.* (2006), Elliott *et al.* (2005), Perlin (2009) and Broussard & Vaihekoski (2012). In a recent paper, Khandani & Lo (2007) discuss the performance of the Lo-MacKinlay contrarian strategies in the context of the liquidity crisis of 2007. These strategies have several common features with the ones developed in this paper. Khandani & Lo (2007) market-neutrality is enforced by ranking stock returns by quantiles and trading "winners-versus-losers", in a dollar-neutral fashion. On the parametric side, Poterba & Summers (1988) study mean-reversion using auto-regressive models in the context of international equity markets. Zebedee & Kasch-Haroutounian (2009) analyzes the impact of pairs-trading at the microstructure level within the airline industry. Avellaneda & Lee (2010) use Principal Component Analysis or sector ETFs in their statistical arbitrage strategy. In all these cases, they model the residuals or idiosyncratic components of a portfolio of pairs, as mean-reverting processes.

In this paper, we investigate the risk and return of a portfolio consisting of many pairs trades all selected based on cointegration. Different from

other authors who used the methodology proposed by Gatev *et al.* (2006) (for example Nath, 2006, Perlin, 2009) and market professionals who have used bollinger bands, in this paper we employ the methodology of cointegration to develop a pairs trading strategy.

The sample period used starts in January 2005 and ends in October 2012, summing up to 1.992 observations. Daily equity closing prices are obtained from Bloomberg. The analysis covers all stocks in the Bovespa index (Ibovespa) from the Sao Paulo stock exchange. An analysis based on Brazilian data is important not only because Bovespa is the largest stock exchange in South America and one of the largest among all emerging economies, but also because the cointegration approach to select pairs have not yet been studied in detail in Brazil. The proposed statistical arbitrage strategy generates average excess return of 16.38% per year in out-of-sample simulations, Sharpe Ratio of 1.34, low exposure to the equity markets and relatively low volatility. The results show the pairs trading strategy based on cointegration is persistently profitable even in the period of global crises, reinforcing the usefulness of cointegration in quantitative strategies.

The remainder of this paper is organized as follows. In section 2, the concepts of statistical arbitrage and pairs trading strategies are presented in greater detail. Section 3 explain the use of cointegration within this class of strategies. In section 4, we describe the strategy proposed. In section 5 the data are discussed and the results obtained from the out-of-sample simulations are empirically verified. In section 6, a conclusion based on the empirical results is presented, along with suggestions of future research.

2. Statistical Arbitrage and Pairs Trading

Statistical arbitrage is a trading or investment strategy used to exploit financial markets that are out of equilibrium. Litterman (2003) explains the philosophy of Goldman Sachs Asset Management as one of assuming that while markets may not be in equilibrium, over time they move to an equilibrium, and the trader has an interest to take maximum advantage from deviations from equilibrium.

Pairs-trading, which is a statistical arbitrage strategy, was pioneered by Nunzio Tartaglias quant group at Morgan Stanley in the 1980's, and it remains an important statistical arbitrage technique used by hedge funds. Tartaglias' group found that certain securities were correlated in their day-to-day price movements, (see Vidyamurthy, 2004). Based on these empiri-

cal investigations, trading strategies might be formed to explore the inefficiencies of stock markets. The key references in this area are Lo & MacKinlay (1988), Khandani & Lo (2007), Lo & MacKinlay (1997), Gatev *et al.* (2006) and Guidolin *et al.* (2009). One of the many possible statistical arbitrage strategies is the pairs trading. In pairs trading we do not deal with trends established for particular assets but with common long-run equilibrium trends among pairs of stocks.

The idea behind pairs trading is to first identify a pair of stocks with similar historical price movement. Then, whenever there is sufficient divergence between the prices in the pair, a long-short position is simultaneously established to bet that the pair's divergence is temporary and that it will converge over time.

Tartaglia and his group used the pairs trading strategy with great success throughout 1987. However, the group was dismantled in 1989, after two years of bad results. Nevertheless, the pairs trading strategy became increasingly popular among individual traders, institutional investors and hedge-funds.

Recently, due to the financial market crisis, it was widely reported in the specialized media that the year 2007 was especially challenging for quantitative hedge funds (see Khandani & Lo, 2007, Avellaneda & Lee, 2010), in particular for the statistical arbitrage strategies. The strategy proposed here is analyzed in the period in question and the results found corroborate those of other authors.

Jacobs & Levy (1993) argues that long-short stock strategy can be implemented in three different ways: as market neutral, as equitizing, or as hedge strategies. Market neutral long-short strategies, as the one proposed here, maintain even exposure to market risks using long and short positions at all times. This approach eliminates exposure to directional risk from the market, such that the obtained return should not be correlated with the market reference index, which is the equivalent to a beta-zero portfolio. The portfolio returns are generated by the isolation of the alfa, adjusted by risk. According to Fung & Hsieh (1999), a strategy is said to be market neutral if its return are independent from the market's relative return. Market neutral funds actively seek to avoid systematic risk factors, betting on relative price movements.

The process of asset pricing can be seen in absolute or relative terms. In absolute terms, asset pricing is made by way of fundamentals, such as discounted future cash flow, for example. Relative pricing means that prices

from assets that are close substitutes for each other should sell for similar prices – it doesn't say what the price of an asset should be, but it says what the price of an asset should be with respect to the price of another asset.

The other pilar of pairs trading is deviations from this relative price. To be able to profit from the trade, these deviations have to be mean reverting. However, reversion to the mean requires a driving mechanism; pairs trading would not work if prices were truly random. The Law of One Price (LOP) is the proposition that two investments with the same payoff in every state of nature must have the same current value. Thus, the price spread between close substitute assets should have a long term stable equilibrium over time. Hendry & Juselius (2001), use this idea to show that short term deviations from these equivalent pricing conditions can create short-lived arbitrage opportunities depending on the duration of price deviation.

Statistical arbitrage is based on the assumption that the patterns observed in the past are going to be repeated in the future. This is in opposition to the fundamental investment strategy that explores and tries to predict the behaviour of economic forces that influence the share prices. Thus statistical arbitrage is a purely statistical approach designed to exploit equity market inefficiencies defined as the deviation from the long-term equilibrium across the stock prices observed in the past.

When a deviation in the spread of the long term equilibrium price relationship is identified to be substantially greater than the slipage[1] due to the bid-ask spread, a position is opened simultaneously, buying the relatively undervalued stock and selling the relatively overvalued stock. The position is closed when the prices return to the spread level of long term equilibrium. The net profit of the operation is the sum of the profits from the long and short positions, calculated as the difference between the open prices and closed prices (ignoring transaction costs).

The natural extrapolation of pairs trading strategies consists of the operation of a group of stocks against another group of stocks, or generalized pairs trading (see Alexander & Dimitriu, 2005a, Dunis & Ho, 2005, Avellaneda & Lee, 2010, Caldeira & Portugal, 2010).

3. Cointegration-based Strategies

The applicability of the cointegration technique to asset allocation was pioneered by Lucas (1997) and Alexander (1999). Its key characteristics,

---

[1]Slipage is defined as the difference between the prices that trigger the order and the prices at which the order is executed.

i.e. mean reverting tracking error, enhanced weight stability and better use of the information comprised in the stock prices, allow a flexible design of various funded and self-financing trading strategies, from index and enhanced index tracking, to long-short market neutral and alpha transfer techniques.

3.1 The cointegration approach

The approach described in Vidyamurthy (2004) is an attempt to parameterize pairs trading strategies exploring the possibility of cointegration. Cointegration is a statistical feature, where two time series that are integrated of order 1, $I(1)$, can be linearly combined to produce one time series which is stationary, or $I(0)$. The pairs trading technique used here is based on the assumption that a linear combination of prices reverts to a long-run equilibrium and a trading rule can be constructed to exploit the expected temporary deviations.

In general, linear combination of non-stationary time series are also non-stationary, thus not all possible pairs of stocks cointegrate.

Definition. A $n \times 1$ time series vector $\mathbf{y}_t$ is cointegrated [2] if

- each of its elements individually are non-stationary and

- there exists a non zero vector γ such that $\gamma \mathbf{y}_t$ is stationary.

In the previous decade the concept of cointegration was increasingly applied in financial econometrics (see Alexander & Dimitriu, 2002). It is an extremely powerful technique, which allows dynamic modelling of non-stationary time-series. The fundamental observation that justifies the application of the concept of cointegration in the analysis of stock prices is that a system involving non-stationary stock prices in levels can have a common stochastic trend (see Stock & Watson, 1988). When compared to the concept of correlation, the main advantage of cointegration is that it enables the use of the information contained in the levels of financial variables. Alexander & Dimitriu (2005a,b), Gatev *et al.* (2006), Caldeira & Portugal (2010), suggest that cointegration methodology offers a more adequate structure for financial arbitrage strategies.

The crucial steps in building the pairs trading strategy is the local estimation of both current and expected spreads. In the framework of cointegration analysis spread is modeled as the local deviation from the long-term

---

[2]For more details about cointegration analysis, see Johansen (1995), Hamilton (1994).

equilibrium among the time series. Therefore the current spread between the assets is computed as the product of cointegrating vector and current stock prices. On the other hand, the expected spread is estimated as the product of cointegrating vector and predicted stock prices. The spread prediction is based on the assumption of a cointegration relation among the pairs of assets.

The idea of pairs trading is to invests an equal amount in asset l and asset s, $\alpha P_t^l = P_t^s$, making this a cashless investment. This can be done by borrowing a number of shares of assets s, immediately selling these and investing the amount in α shares of asset l. Thus, we define the logarithm of the investment equation as follows:

$$0 = \log(\alpha) + \log(P_t^l) - \log(P_t^s). \tag{1}$$

The minus sign reflects the fact that asset s is sold short. The log-return on this investment over a small horizon $(t - 1, t)$ is

$$\log\left(\frac{P_t^l}{P_{t-1}^l}\right) - \log\left(\frac{P_t^s}{P_{t-1}^s}\right). \tag{2}$$

Thus, to make profit the investor doesn't need to predict the behavior of P_t^l and P_t^s, but only that of the difference $\ln(P_t^l) - \ln(P_t^s)$. If we assume that $\{\ln(P_t^l), \ln(P_t^s)\}$ in (1) is a non-stationary $\text{VAR}(p)$ process, and there exists a value γ such that $\ln(P_t^l) - \gamma \ln(P_t^s)$ is stationary, we will have a cointegrated pair. The investment equation will then become

$$0 = \ln(\alpha) + \ln(P_t^l) - \gamma \ln(P_t^s). \tag{3}$$

The value of γ will be determined by cointegration, and the long run equilibrium relationship between the assets determines α. The return on the investment will be

$$\ln\left(\frac{P_t^l}{P_{t-1}^l}\right) - \gamma \ln\left(\frac{P_t^s}{P_{t-1}^s}\right). \tag{4}$$

If $\gamma = 1$, the investor is able to profit from the trade, even though the investment has an initial value of 0. A γ close to zero requires funds to invest in l. A large γ exposes the investor to risk of going short on s.

Nevertheless, the cointegration approach to pairs trading has its limitations. These limitations are mainly due to two implementation problems:

poor estimates of the cointegrating vector; and inaccurate prediction of the expected spread that can generate false trading signals. Moreover, the 2-step cointegration test procedure renders results sensitive to the ordering of the variables, therefore the residuals may have different statistical properties. Additionally, if the bivariate series are not cointegrated, the "cointegrating regression" leads to spurious estimates (Lim & Martin, 1995) , making the mean reversion analysis on residuals unreliable. So what can be done to improve this simple but intuitive approach? One way is to perform more rigorous testing of cointegration, including using Johansen's testing approach based on a Vector Error Correction Model (VECM) and comparing the outcome to the Engle-Granger results. More importantly, if the cointegration test fails, one should refrain from trading based on residuals whose properties are unknown. However, the ultimate test of the cointegration approach to pairs trading will be given by the out-of-sample profitability of the strategy.

3.2   The Model

The investment strategy we aim at implementing is market neutral, thus we will hold a long and a short position both having the same value in local currency. This approach eliminates net equity market exposure, so the returns provided should not be affected by the market's direction.

Typically the pairs trading algorithm has two main parts. The first fundamental building block of this methodology is a pairs selection algorithm which, in our case, is essentially based on cointegration testing. The objective of this phase is to identify pairs whose linear combination exhibits a significant predictable component that is uncorrelated with underlying movements in the market as a whole. With this aim, we first check if all the series are integrated of the same order, $I(1)$. This is done by way of the Augmented Dickey Fuller Test (ADF). Having passed the ADF test, cointegration tests are performed on all possible combination of pairs. To test for cointegration we adopt Engle and Grangers 2-step approach and Johansen test.[3]

For detected cointegrating relations, the second part of the algorithm creates trading signals based on predefined investment decision rules. In order to implement the strategy we need to follow a couple of trading rules, i.e. to determine when to open and when to close a position. First, we

---

[3]All of the procedures are implemented on MATLAB software, version 7.0. The functions cadf and johansen are used, and are available at www.spatial-econometrics.com.

calculate the spread between the shares. The spread is calculated as $\varepsilon_t = P_t^l - \gamma P_t^s$, where ε_t is the value of the spread at time t. Accordingly, we compute the dimensionless z-score defined as $z_t = \frac{\varepsilon_t - \mu_\varepsilon}{\sigma_\varepsilon}$, the z-score measures the distance to the long-term mean in units of long-term standard deviation.

Our basic rule will be to open a position when the z-score hits the 2 standard deviation thresholds from above or from bellow. This situation implies that the stocks are mispriced in terms of their relative value to each other. If the z-score hits the -2 standard deviation threshold, it means that the portfolio of pairs is bellow its long-run equilibrium value. In this case, one should buy the portfolio, which means buying stock l and selling stock s. If the z-score hits the 2 standard deviation threshold from above, the portfolio of pairs is overvalued and one should sell it short, which means selling stock l and buying stock s. The position is closed the z-score approaches zero again. In all cases opening or closing a position means buying and selling the stocks simultaneously. More specifically, the basic trading signal can be summarized as

$$
\begin{aligned}
\text{Buy to open if} \quad & z_t < -2.00 \\
\text{Sell to open if} \quad & z_t > 2.00 \\
\text{Close short position if} \quad & z_t < 0.75 \\
\text{Close long position if} \quad & z_t > -0.50
\end{aligned}
$$

The rationale for opening trades only when the z-score z_i is far from equilibrium is to trade only when we think that we detected an anomalous excursion of the co-integration residual.[4] As in Avellaneda & Lee (2010), the threshold values to open and to close positions are selected empirically. Furthermore, there will be some additional rules to prevent us from loosing too much money on one single trade. If the ratio develops in an unfavourable way, we will use a stop-loss and close the position if we have a loss of 7%. Stop loss constraints are not always considered in academic research (for example, Elliott *et al.*, 2005, Gatev *et al.*, 2006, Perlin, 2009,

---

[4]Since we deal with many pairs trades simultaneously, there is the possibility of having a trade where stock A is sold short, and another trade where we buy stock A. In this rare event, we perform both operations. Although it increases transaction costs, this approach is computationally easier to implement.

Gatarek *et al.*, 2011), an exception is Nath (2006) that adopts a stop-loss trigger to close the position whenever the distance widens further to hit the 5th or the 95th percentile. However, stop loss constraints are fundamental in practice to avoid large losses.[5] Finally, we will never keep a position for more that 50 days, since in-sample profitability of the strategy decreases with time (see Figure 1). Based on our in-sample results, 50 days should be enough time for the pairs to revert to equilibrium, but also a short enough time not to loose time value. On average, the mean reversion will occur in approximately 10 days, and there is no reason to wait for a pair to revert fully, if there is very little return to be earned. The rules described are totally based on statistics and predetermined numbers.
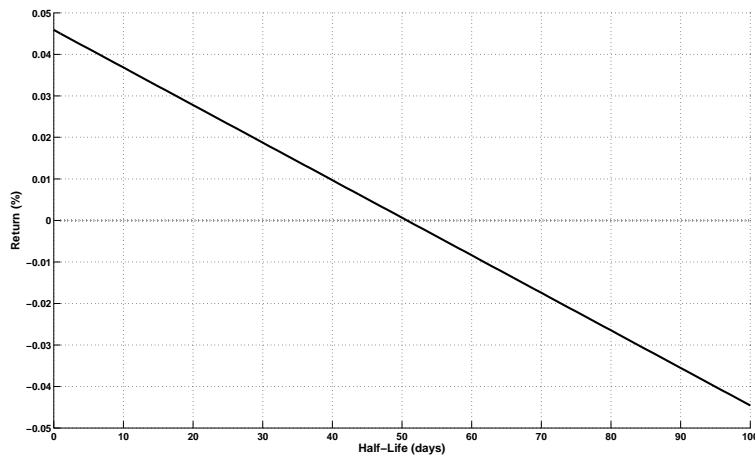
Figure 1

Pairs Trading Profitability and Speed of Convergence

The lines plots mean returns versus half-life from the pairs trading strategy. The pairs trading strategy involves matching stock pairs based on cointegrated prices over a one year estimation period. Then, during the following four months, the strategy looks for instances in which the z-score of the two stocks in the pair diverge by more than two standard deviations. When there is divergence the strategy buys the stock that went up and shorts the stock that went down.

As an additional criteria for selecting the pairs to be used in the composition of the portfolio, we apply an approach introduced by Dunis *et al.* (2010). The approach consists in selecting the pairs for trading based on

---

[5]We have consider stop loss constraint of 3%, 5% and 7%. The results are similar, however, when a position looses 7% it rarely comes back to a positive performance.

the best in-sample sharpe ratios.[6] We form the portfolio of 20 best trading pairs that present the greatest SR in the in-sample simulations and use them to compose a pairs trading portfolio to be employed out-of-sample. Goetzmann *et al.* (2002) and Gatev *et al.* (2006) show that Sharpe Ratios can be misleading when return distributions have negative skewness. This is unlikely to be a concern for our study, since our Table 1 showed that the returns to pairs portfolios are positively skewed, which – if anything – would bias our Sharpe Ratios downward. The portfolio is expected to produce a positive return as valuations converge.

As soon as the spread distances itself from its long term mean, one can bet that the spread will return to its long term mean, however we do not know if we will gain more on long or on short positions.[7] Once a trade is initiated, the portfolio is not rebalanced. Therefore, after the opening of a position, even when prices move and the position may no longer be neutral, the portfolio is not rebalanced. We only have two types of transactions that are admitted by the strategy's methodology; move into a new position, or the total liquidation of a previously opened position. The strategy adopted here seeks to be beta-neutral, thus the financial values allocated to long and short stocks might not be equal.[8]

4. Data and Empirical Results

4.1 Data

The data used in this study consists of daily closing prices of the 50 stocks with largest weights in the Ibovespa index from Sao Paulo Stock Exchange in the beginning of each trading period, which lasts four months. All of the stocks used are listed in Bovespa, which means they are among the most liquid stocks traded on the Brazilian market. This characteristic is important for pairs trading, since it often diminishes the slipage effect. Moreover, using less liquid stocks may involve greater operational costs (bid and ask spread) and difficulty in renting a stock. Since the stocks in Ibovespa change every 4 months, stocks in the sample are also changed,

---

[6]Sharpe ratio is calculated as the ratio of annualized return to annualized standard deviation.

[7]We don't know which case occurs first: if the stocks return to their long term equilibrium because the overvalued stock price falls more or because the undervalued stock price climbs more, or if they both present the same performance.

[8]One alternative is to define the investment as financially-neutral, assuming equal financial volumes of long and short shares.

thus the sample is not subject to survivor bias. The data were obtained from Bloomberg, taken from the period of January 2005 to October 2012. The data are adjusted for dividends and splits, avoiding false trading signals generated by these events, as documented by Broussard & Vaihekoski (2012).

It is common in pairs trading strategies to require that the stock pairs belong to the same sector, for example in Chan (2009) and Dunis *et al.* (2010). Here, we do not adopt this restriction. Therefore, stock pairs from companies belonging to different sectors can be traded, as long as they satisfy the cointegration criterion.

4.2   Estimation and Out-of-sample Results

Initially, we divide the sample into training, and testing periods. The training period is a preselected period where the parameters of the experiment are computed. Immediately after the training period, the testing period follows, where we run the experiments using the parameters computed in the first period. Note that pairs are also treated as parameters in our trading system. We use one year for training and four months for testing.

Cointegration tests are performed (Johansen and Engle-Granger) for all possible combinations. Of the 1,225 possible pairs, an average of 94 cointegrated pairs were obtained in each period. The pairs that passed the cointegration tests are then ranked based on the in-sample SR,[9] following Gatev *et al.* (2006) and Andrade *et al.* (2005). After selecting 20 pairs with highest SR, four months of pairs trading are carried out. At the end of each trading period the position that were opened are closed, and a new training period ending on the last observation of the previous trading period is initiated. Now stocks and pairs can be substituted and all parameters are re-estimated. This procedure continues in a rolling window fashion until the end of the sample.

We indicate the price of the stock in which we have a long position on day t as P_t^l and the price of the stock we are shorting on day t as P_t^s. Thus, the net return for pair i on day t can be defined as,

$$r_{it}^{raw} = \ln\left[\frac{P_t^l}{P_{t-1}^l}\right] - \gamma \ln\left[\frac{P_t^s}{P_{t-1}^s}\right] + 2\ln\left(\frac{1-C}{1+C}\right) \tag{5}$$

---

[9]The trading gains and losses are computed over long-short positions of one BRL, the payoffs have the interpretation of excess returns.

where $w_{j,t}$ is the portfolio weight for pair j at time t, N is the number of pairs considered, and C refers to transaction costs. The second part of Equation (5) has the objective of accounting for transaction costs.[10]

The simple net return of a portfolio consisting of N pairs (or assets) is a weighted average of the simple net returns of the pairs involved, where the weight on each pair is the percentage of the portfolio's value invested in that pair. Let p be a portfolio that places weight w_i on pair i. Then the simple return of p at time t is $R_t^p = \sum_{i=1}^{N} w_{it} R_{it}$, where R_{it} is the simple return of pair i. The log-returns (or continuously compounded returns) of a portfolio, however, do not have the above convenient property. Notice that,

$$r_t^p = \ln(1 + R_t^p) = \ln \left(1 + \sum_{i=1}^{N} w_{it} R_{it} \right) \neq \sum_{i=1}^{N} w_{it} r_{it}$$

where r_{it} denotes the one-period log-return on pair i as in (5). When returns are measured over short intervals of time, and are all small in magnitude, the continuously compound return (log-return) on a portfolio is close to the weighted average of the continuously compounded returns on the individual assets: $R_t^p \approx \sum_{i=1}^{N} w_{it} r_{it}$. Thus, as pointed out by Campbell *et al.* (1997) and Tsay (2010), this approximation is often used to study portfolio returns. Nevertheless, here we transform back the log-returns of each pair into simple returns in order to accurately compute the return of the portfolio of pairs.

Let R_{it} denote the simple daily return on pairs trading strategy. The continuously compounded monthly return, r_{it}, is defined as:

$$r_{it} = \ln(1 + R_{it}) = \ln \left(\frac{P_t}{P_{t-1}} \right),$$

Given a monthly continuously compounded return r_{it}, it is straightforward

---

[10]This formula accounting directly for transaction costs can be explained in a intuitive way. Suppose we buy the stock ξ at P_{t-1}^{ξ} at time $t-1$ and to sell it one step after, at time t, at P_t^{ξ} unit of money (price quotation), in fact the costs of buying are $P_{t-1}^{\xi}(1+C)$ and the profit coming from selling $P_t^{\xi}(1-C)$. This corresponds to the decomposed net return:

$$\ln \left[\frac{P_t^{\xi}(1-C)}{P_{t-1}^{\xi}(1+C)} \right] = \ln \left[\frac{P_t^{\xi}}{P_{t-1}^{\xi}} \right] + \ln \left[\frac{(1-C)}{(1+C)} \right] = r_t^{\xi} + \ln \left[\frac{(1-C)}{(1+C)} \right]$$

where the second term accounts for the transaction costs of buying and selling. See, for example, Perlin (2009).

to solve back for the corresponding simple net return R_{it}:

$$e^{r_{it}} = 1 + R_{it} \Longrightarrow R_{it} = e^{r_{it}} - 1.$$

The daily net return to a portfolio of N pairs on day t is

$$R_t^{net} = \sum_{i=1}^{N} w_{it} R_{it}, \tag{6}$$

where w_{it} is the weight of each pair in the portfolio, which in our application is $1/N$.

We consider transaction costs of 0.5% one-way in total for both shares following Dunis *et al.* (2010), Dunis & Ho (2005) and Alexander & Dimitriu (2002). We are dealing with the 50 most liquid Brazilian shares in this paper. Transaction costs consist of 0.1% of brokerage fee for each share (thus 0.2% for both shares), plus slippage for each share (long and short) which we assume to be 0.05% (see Avellaneda & Lee, 2010), and 0.2% of rental cost for short positions (average rental cost is 2% per year per share). In this paper we use a fully invested weighting scheme. As argued by Broussard & Vaihekoski (2012), the fully invested scheme is less conservative as it assumes capital is always divided between the pairs that are open. In practice, we assume that each pair is given the same weight at the beginning of the trading period. If a pair is not opened, the return is zero, and thus the weight does not change. This implies that for the fully invested weighting scheme, the money from a closed pair is invested in the other pairs that are open. If a pair is reopened, the money is invested back by redistributing the investment between the pairs according to their relative weights.

We examine the pairs trading portfolios performance in terms of the cumulative return, variance of returns ($\hat{\sigma}^2$), Sharpe ratio (SR) and Maximum Drawdown (MDD). The Drawdown is the measure of the decline from a historical peak in some variable (typically the cumulative profit or total open equity of a financial trading strategy). In this paper, MDD is defined as the maximum percentage drop incurred from a peak to a bottom in a certain time period. These statistics are computed as follows:

$$R^A = 252 * \frac{1}{T} \sum_{t=1}^{T} R_t$$

$$\hat{\sigma}^A = \sqrt{252} * \frac{1}{T} \sum_{t=1}^{T} (R_t - \hat{\mu})^2$$

$$SR = \frac{\hat{\mu}}{\hat{\sigma}}, \text{ where } \hat{\mu} = \frac{1}{T} \sum_{t=1}^{T} w_{it} R_{it}$$

$$MDD = \sup_{t \in [0,T]} \left[\sup_{s \in [0,t]} R_s - R_t \right]$$

By construction, the maximum Drawdown is higher in absolute values than the maximum loss. Indeed, this is the most pessimistic scenario.

Table 1 shows the pairs used in the strategy during the last quarter of 2009. We find out that most pairs are independent random walk. However, of the 1,225 possible pairs, 97 passed the cointegration tests of Johansen and Engle-Granger. Of those 97, 20 pairs that presented the greatest in-sample SR were selected to be used out-of-sample. Even though there weren't restrictions requiring stocks within a pair to come from companies from the same sector, the majority of pairs are comprised of stocks from companies that are in some way related. One also notes that many of the pairs present a half life of less than 10 days, reinforcing the mean-reversion characteristic, which is desirable for the strategy. Although all pairs present positive SR in-sample simulations, not all obtained positive return in the out-of-sample trading period. During the period in question, 6 of the 20 pairs that comprise the portfolio showed negative results, and on average a net return of 3.82% per pair was obtained.
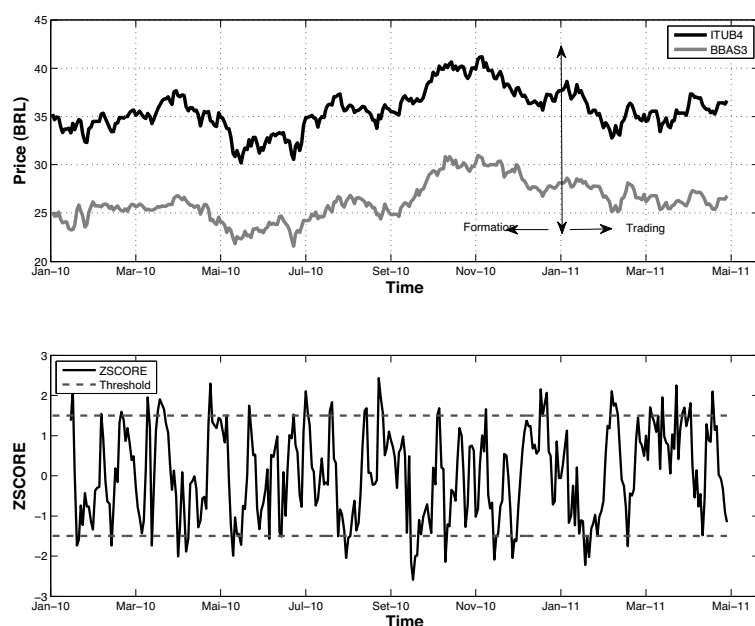
Table 1
Descriptive atatistics of the pairs. Sample period 2009:09 to 2009:12

	Stock 1	Stock 2	EG (ADF)	JH (λ_{tr})	SR(*in-sample*)	Half-Life	Net Ret.
	Itub4	Itsa4	-3.67	18.32	4.33	4.58	6.60
	Usim5	Usim3	-3.47	18.00	3.29	16.82	9.03
	Vale3	Brap4	-4.25	24.23	3.13	12.19	7.96
	Ambv4	Natu3	-4.27	19.42	2.69	6.00	10.19
	Ambv4	Jbss3	-4.35	20.63	2.57	9.01	4.00
	Csna3	Brap4	-3.78	19.27	2.39	11.62	10.10
	Bbas3	Lren3	-3.47	16.86	1.90	6.31	-19.68
	Cyre3	Gfsa3	-4.17	18.20	1.70	4.81	24.19
	Vale3	Ccro3	-3.39	14.24	1.61	10.87	-4.00
Pairs	Bbas3	Usim3	-3.69	19.58	1.60	8.61	-10.69
	Brfs3	Pcar5	-4.15	19.20	1.59	6.41	-6.59
	Netc4	Jbss3	-3.87	17.23	1.58	8.39	-8.28
	Cple6	Pcar5	-4.05	17.26	1.57	12.22	13.78
	Cple6	Ccro3	-3.58	15.97	1.56	13.25	5.37
	Lame4	Ambv4	-3.85	17.06	1.46	14.85	8.91
	Itub4	Ccro3	-3.92	17.60	1.45	6.18	13.45
	Usim5	Bbas3	-4.39	22.13	1.40	6.51	0.40
	Ambv4	Ccro3	-3.82	18.76	1.24	6.98	22.09
	Brfs3	Cple6	-3.55	19.38	1.15	7.04	-14.38
	Bvmf3	Netc4	-3.98	19.39	1.10	17.01	4.02

Note: half-life is expressed in days and net return in %. The 95% critical values for Johansen test is 13.43 and for ADF is 3.38. "EG" (ADF) refers to Engle-Granger cointegration test (Augmented Dickey Fuller); "JH" refers to Johansen cointegration test; "SR" refers to Sharpe Ratio for in-sample period; "Half-Life" is the expected time to revert half of its deviation from the mean measured in trading days and "Net Ret" refers to net return for out-of-sample period.

See Figure 2 for a graph showing the evolution of the z-score of residuals of ITUB4 against BBAS3. The z-score measures the distance to equilibrium of the cointegrated residual in units of standard deviations, i.e. how far away a given pair is from the theoretical equilibrium value associated with our model.

Figure 2

Evolution of the stock prices and z-score of ITUB4 versus BBAS3 from Sep 2008 to Jan 2010

Note: Normalized spread and the times when the positions are open. The pair is open every time its spread exceeds the thresholds.

Table 2 summarizes the excess returns for the pairs portfolios. The results presented refer to the out-of-sample analysis (from January, 2006 to October, 2012). The profitability shown has already been discounted for transaction costs.[11] One can also note that the strategy presents a relatively low volatility of 12.49% in annualized terms, and a correlation coefficient with the market of -0.103, indicating that the strategy can be considered market neutral.

---

[11]The costs considered are 0.5% in opening and 0.5% in closing the position, summing up to 0.10% per operation. Costs related to renting stocks sold short were considered to be 2% per year.

Table 2
Statistics of excess returns of unrestricted pairs trading strategies, 2006:01 to 2012:10

Summary Statistics of the Pairs Trading Strategy	
# of observations in the sample	1992
# of days in the training window	250
# of days in the trading period	84
# of trading periods	21
# of pairs in each trading period	20
# min of cointegrated pairs in a trading period	53
# max of cointegrated pairs in a trading period	195
Average annualized return	16.39%
Annualized volatility	12.42%
Annualized Sharpe Ratio	1.34
Largest daily return	6.25%
Lowest daily return	-4.80%
Cumulative profit	189.29%
Spearman correlation coefficient	0.076
Skewness	0.41
Kurtosis	5.62
Maximum *Drawdown*	24.49%

Note: Summary statistics of the daily and annual excess returns on portfolios of pairs between Jan 2006 and October 2012 (1 and 250 observations, respectively). The sample period is from January 2005 through October 2012, while the out-of-sample simulations were performed from January 2006 through October 2012. In particular, we report the minimum, median, mean, the skewness and kurtosis, and the maximum of three important performance measures: the accumulative profit, the Sharpe Ratio, and the maximum draw down.

We also present the maximum drawdown of the strategy in the analyzed period, which was 24.49%. This is a simple measure of the fall in percentage terms with respect to the peak of the cumulative return, and can be used as a measurement of how aggressively the strategy's leverage can be increased. It can be seen in Figure 3, which conveys the strategy's cumulative profit and volatility, that the maximum drawdown occurred in the first semester of 2008.

Figure 3 compares the cumulative excess returns and volatility of the strategy with the cumulative excess returns of the Ibovespa index. The smooth path of the pairs trading portfolio contrasts dramatically with the volatility of the stock market. It can be noted in the second panel of Figure 3, that the pairs trading strategy presented a relatively low and stable standard of volatility for practically the entire analyzed period, running at levels below 15% in annualized terms, for nearly the entire period. Even in the most acute period of the international financial crisis, when the volatility

on the domestic stock market surpassed 120%, the volatility of the strategy never reached 40%.
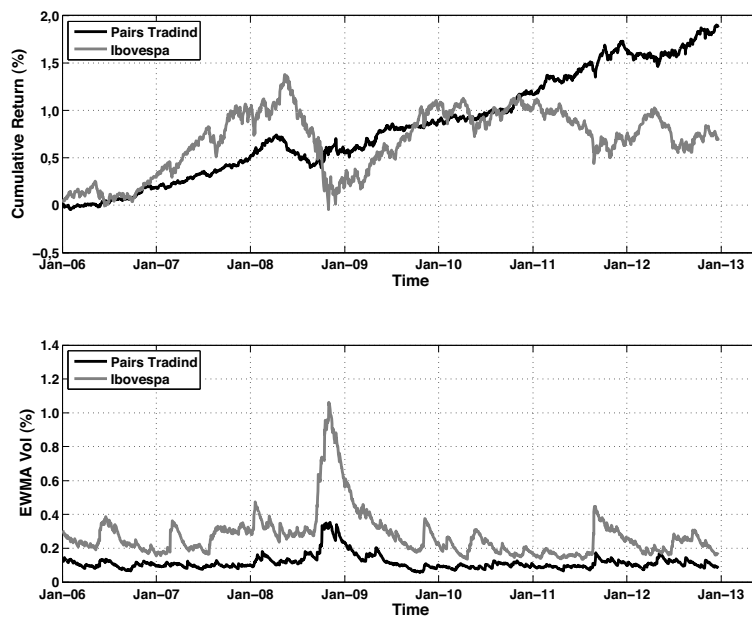
Figure 3

Cumulative excess return and Volatility of top 20 pairs and Ibovespa. Period 2006:01 – 2012:10
Note: In the first panel, cumulative profit of the pairs trading strategy and Bovespa, in the second annualized volatilities (EWMA Vol with $\lambda = 0.94$).

Another relevant part of the evaluation of the pairs trading strategy is the analysis of the correlation with the main reference market index, since one of the goals of the strategy is market neutrality (see Alexander & Dimitriu, 2002). The strategy showed a correlation with Ibovespa of less than 0.15, and for a good part of the sample it was less than 0.10, as can be observed in Figure 4. The estimated market β of the portfolio,[12] is also presented, which corroborates with the strategy's market neutrality.

---

[12]The market β was estimated with the Kalman filter with the goal of verifying its stability over time.
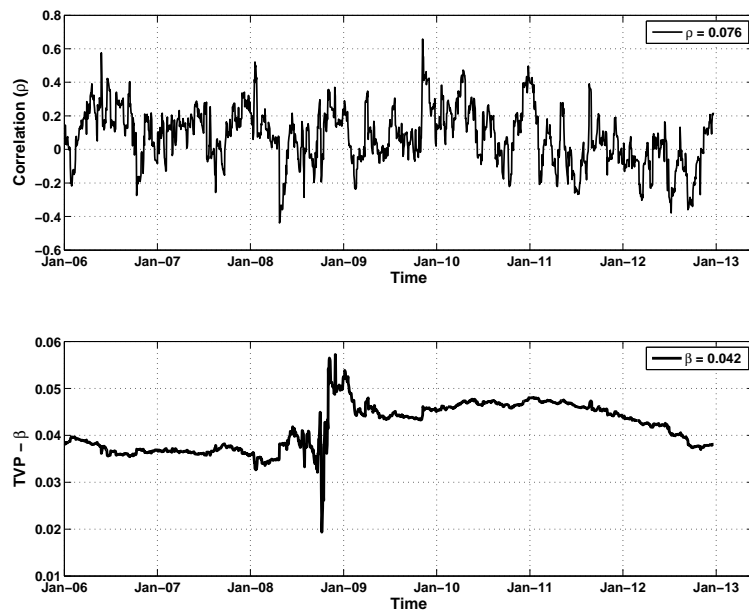
Figure 4
Spearman correlation coefficient and β of the Strategy
The Spearman correlation coefficient calculated based on a sliding window of 84 observations. β estimated by the Kalman filter.

Table 3 summarizes annual statistics of pairs trading strategies. The tables include mean, median, standard deviation, skewness, kurtosis, Sharpe Ratio and Maximum drawdown. To test the statistical significance of the difference between the variances and Sharpe ratios of the returns for pairs trading and Ibovespa, we follow DeMiguel *et al.* (2009) and use the stationary bootstrap of Politis & Romano (1994). The p-values reported on Table 3 are computed using the stationary bootstrap of Politis & Romano (1994) generating 1000 bootstrap samples with smoothing parameter q = 0.25.

It can be observed in Table 3 that the strategy showed its worst performance in the year 2008, accumulating a net profit of 2.85% and volatility slightly higher than in other years. As highlighted by Khandani & Lo (2007) and Avellaneda & Lee (2010), the second semester of 2007 and first semester of 2008 were quite complicated for quantitative investment funds. Particularly for statistical arbitrage strategies that experienced significant losses during the period, with subsequent recovery in some cases. Many

managers suffered losses and had to deleverage their portfolios, not bene-fiting from the subsequent recovery. We obtain results which are consistent with Khandani & Lo (2007) and Avellaneda & Lee (2010) and validate their unwinding theory for the quant fund drawdown. Note that in Figure 3, the proposed pairs trading strategy presented significant losses in the first semester of 2008, starting its recovery in the second semester. Khandani & Lo (2007) and Avellaneda & Lee (2010) suggest that the events of 2007-2008 may be a consequence of a lack of liquidity, caused by funds that had to undo their positions.

Table 3
The P&L (in %) of the statistical arbitrage strategy for 7 years

Year	Max	Min	Median	Mean	Std	Skew	Kurt	Accum	Sharpe	MDD
2006	4.96	-4.54	0.137	0.127	0.103	0.040	3.48	18.28	1.72	5.57
					(0.0000)			(0.0170)	(0.0000)	
2007	4.95	-6.63	0.307	0.186	0.098	0.277	3.67	28.59	2.69	4.14
					(0.0000)			(0.0000)	(0.0000)	
2008	14.66	-11.39	0.012	-0.151	0.199	0.674	6.76	1.58	0.12	24.98
					(0.0000)			(0.4880)	(0.4520)	
2009	7.17	5.24	0.279	0.239	0.116	0.34	5.24	22.85	1.80	5.24
					(0.0000)			(0.0300)	(0.0000)	
2010	4.10	-4.73	0.027	0.003	0.097	0.07	3.66	14.17	1.49	3.66
					(0.0000)			(0.0260)	(0.0000)	
2011	5.10	-8.085	0.000	-0.065	0.105	0.31	4.18	17.51	1.35	8.34
					(0.0000)			(0.0018)	(0.0000)	
2012	4.72	-3.31	-0.003	0.017	0.113	-0.28	4.47	13.62	1.38	5.76
					(0.0000)			(0.0000)	(0.0000)	
All Time	14.66	-11.39	0.081	0.050	0.119	0.439	8.809	166.97	1.241	24.98

Note: Summary statistics for the annual percentage excess (net) returns on portfolios of top 20 pairs between Jan-2006 and Out-2012. The P-values are computed using the stationary bootstrap generating 1000 bootstrap samples with smoothing parameter $q = 0.25$. The P-values for $q = 0.75$ and 0.50 are similar (available upon request), which is consistent with White (2000). The bootstrap reality check p-value is zero for all years except for 2008, which indicates that the average return is not the result of data snooping.

4.3 Bootstrap method for assessing pairs trading performance

The performance measurement of a trading strategy is then the crucial point to determine it usefulness. To check for the profitability of the strategy, different techniques can be applied. Among all, the simplest is for sure the comparison with a naive trading strategy. This can be done for example by considering a bootstrap method, where we generate randomly trading signals and trade according to them. This exercise is useful to determine if our results could be obtained purely by chance.

In this section we further explore whether our pairs trading strategies are merely a disguised way of exploiting these cointegration relationship.

As Gatev *et al.* (2006), we conduct a bootstrap where we compare the performance of our pairs to random pairs. The starting point of the bootstrap is the set of historical dates in which the various pairs are opened. In each bootstrap we replace the actual stocks with two random securities with similar prior one-day returns as the stocks in the actual pair. Similarity is defined as coming from the same decile of previous day's performance. The difference between the actual and the random pairs returns provides an indication of the portion of our pairs return that is not due to reversion. We bootstrapped the entire set of trading dates 2500 times. On average we find that the returns on the random pairs are well below the return based on cointegrated pairs. In fact, the excess returns of random pairs are slightly negative, and the standard deviations of the returns are large relative to the true pairs, which is a reflection of the fact that the simulated pairs are poorly matched. Considering, as before, the same percentage transaction costs for both the long as the short positions, we finally obtain the net return of the naive strategy:

$$R_t^{Naive} = \sum_{i=1}^{N} w_{i,t} r_{i,t} + 2N \ln \left(\frac{1-C}{1+C} \right) \qquad (7)$$

4.4 White's reality check and Hansen's SPA test

Reality check studies use White (2000) bootstrap reality check methodology to assess data snooping bias associated with an in-sample search for profitable trading rules. White's statistical procedure can directly quantify the effect of data snooping by evaluating the performance of the best trading rule in the context of the full universe of rules. The best trading rule is found by searching over the full set of trading rules and selecting the rule that maximizes a pre-determined performance criterion (e.g. average net return). The p-value for the best trading rule is found by simulating the asymptotic distribution of the maximum of the performance measure across the full universe of trading rules. A reality check p-value for the best trading rule can be considered a data-snooping-adjusted p-value.

Let f_k denote the excess return of the k-th trading rule over a benchmark and $\psi_k = \mathbb{E}[f_k]$ for $k = 1, \ldots, M$. The null hypothesis is that there does not exist a superior model in the collection of M models (rules):

$$\mathbb{H}_0 : \max_{1 \leq k \leq M} \psi_k \leq 0 \qquad (8)$$

Rejecting (8) indicates that there is at least one model that outperforms the benchmark. An obvious test statistic for this hipostis is the maximum of the normalized sample average of $f_{k,i}$, i.e.,

$$\overline{V}_n = \max_{1 \leq k \leq M} \sqrt{n} \overline{f}_k, \qquad (9)$$

where $\overline{f}_k = \sum_{i=1}^n f_{k,i}/n$, with $f_{k,i}$ being the ith observation of f_k, and $f_{k,1}, \ldots, f_{k,n}$ are the computed returns in a sample of n past prices for the kth trading rule. Letting $\{f_{k,1}^*(b), \ldots, f_{k,n}^*(b)\}$ denote the bth bootstrap sample and $\overline{f}_k^*(b) = \sum_{i=1}^n f_{k,i}^*(b)/n$, White (2000) proposed to approximate the sampling distribution of $\max_{1 \leq k \leq M} \sqrt{n}\left(\overline{f}_k - \psi_k\right)$ by the empirical distribution of

$$\overline{V}_n^*(b) = \max_{1 \leq k \leq M} \sqrt{n}\left[\overline{f}_k^*(b) - \overline{f}_k\right], \quad b = 1, \ldots B. \qquad (10)$$

To test (8), he proposed to compute

$$\hat{p} = \left\{\# \text{ of bootstrap samples with } \overline{V}_n^*(b) > \overline{V}_n\right\}/B \qquad (11)$$

and to reject (8) if $\hat{p} < \alpha$, where α is some specified significance level, e.g., 0.05. The rationale underlying this bootstrap test is that $\hat{p}$ is a bootstrap estimate of the p-value of the test statistic (9); i.e., the probability that (10) exceeds its observed value under the extremal point $\psi_1, \ldots, \psi_M = 0$ of the null hypothesis.

Hansen (2005) pointed out two problems with the preceding bootstrap test, which is called *White's reality check*. First, the average returns $\overline{f}_k$ are not studentized. Second, despite the fact that the null hypothesis (8) consists of an infinite number of parameter values, the distribution of White's reality check is based on the least favorable null $\psi_1, \ldots, \psi_M = 0$. To address these problems with White's reality check, Hansen proposed a superior predictive ability (SPA) test that replaces $\sqrt{n}\overline{f}_k$ in (9) by its k studentized version

$$\tilde{V}_n^*(b) = \left(\max_{1 \leq k \leq M} \frac{\sqrt{n}\overline{f}_k}{\hat{\sigma}_k}\right)_+, \qquad (12)$$

where $x_+ \max x, 0$ and $\hat{\sigma}_k$ is a consistent estimator of the standard deviation of $\sqrt{n}\overline{f}_k$. The SPA test also uses a different method to bootstrap the distribution of $\tilde{V}_n$. Defining $Z_{k,i}^*(b) = f_{k,i}^*(b) - \overline{f}_k \cdot \mathbf{1}_{\{\overline{f}_k \geq -n^{-1/4}\hat{\sigma}/4\}}$ and

letting $\overline{Z}_k^*(b)$ denote the sample average of the bth bootstrapped sample $\left\{\overline{Z}_{k,i}^*(b)\right\}_{i=1,\dots,n}$, Hansen (2005) used $\tilde{V}_n$ in place of $\overline{V}_n$ and the empirical distribution of

$$\tilde{V}_n^*(b) = \left(\max_{1 \leq k \leq M} \frac{\sqrt{n}\overline{Z}_k^*(b)}{\hat{\sigma}_k}\right)_+ , \quad b = 1,\dots B, \qquad (13)$$

Results indicate that the cointegrated based pairs trading strategy over 2005–2012 generates an annual mean return of 15.87% (a break-even transaction cost of 0.20% per trade). The bootstrap reality check p-value is zero for all years except for 2008, which indicates that the average return is not the result of data snooping. The results for the Sharpe ratios indicate that the pairs trading strategy based on cointegration deliver a better risk-adjusted performance in comparison to the benchmark model. For instance, the strategy delivers average Sharpe ratio of 1.241 during the out-of-sample period and the differences in SR are statistically significant.

Table 4
Reality Check: Out-of-Sample Strategy Performance

This table reports the out-of-sample (2006-2010) performance of the pairs trading based on cointegration relative to bootstrap algoritm trading based on the mean return criterion and Sharpe ratio. The p-values refers to consistent p-value of Hansen.

Year	Original P-value (White)			Consistent P-value (Hansen)		
	Return	Volatility	Sharpe Ratio	Return	Volatility	Sharpe Ratio
2006	0.0770	0.0000	0.0000	0.0770	0.0000	0.0000
2007	0.0480	0.0000	0.0000	0.0490	0.0000	0.0000
2008	0.3110	0.0000	0.3910	0.3120	0.0000	0.3910
2009	0.1230	0.0000	0.0000	0.1230	0.0000	0.0000
2010	0.0310	0.0000	0.0000	0.0310	0.0000	0.0000
2011	0.0870	0.0000	0.0000	0.088	0.0000	0.0000
2012	0.0220	0.0000	0.0000	0.0220	0.0000	0.0000
All Time	0.0068	0.0000	0.0000	0.0068	0.0000	0.0000

5. Conclusions

In this paper we have proposed a statistical arbitrage strategy known as pairs trading for stocks of Sao Paulo stock exchange. The strategy is implemented based on cointegration, exploring the the mean-reversion of pairs. Cointegration tests are applied to all possible pair combinations in order to identify stock pairs that share a long term equilibrium relationship. Of 1,225 possible pairs, on average, 90 cointegrated pairs from each forma-

tion period were obtained. Subsequently, we calculated the standardized spread between the stocks and we simulated trades in-sample. From there, a diversified portfolio containing 20 pairs that displayed the greatest SR in-sample were selected to be traded out-of-sample.

The cumulative net profit from the four year period of rolling window out-of-sample tests was of 189.29%, with an annual mean of 16.38%. In addition, the pairs trading here implemented showed relatively low levels of volatility and no significant correlation to Ibovespa, confirming its market neutrality. The results are attractive when compared to other strategies employed by hedge funds (see Soerensen, 2006). Specially if we take into account that the strategy is practically cashless. In future research projects we will try to enhance profitability and to mitigate risks through a method to identify the stability of the cointegration parameters. Another goal is to apply the proposed methodology to high frequency data. The results presented reinforce the use of the concept of cointegration as an important tool in the quantitative management of funds.

References

Alexander, Carol. 1999. Optimal hedging using cointegration. *Philosophical Transactions of the Royal Society Series A*, **357**, 2039–2058.

Alexander, Carol, & Dimitriu, Anca. 2002. *The Cointegration Alpha: Enhanced Index Tracking and Long-Short Equity Market Neutral Strategies*. ISMA Discussion Papers in Finance 08. ISMA Finance.

Alexander, Carol, & Dimitriu, Anca. 2005a. Indexing and statistical arbitrage: tracking error or cointegration? *Journal of Portfolio Management*, **31**, 50–63.

Alexander, Carol, & Dimitriu, Anca. 2005b. Indexing, cointegration and equity market regimes. *International Journal of Finance and Economics*, **10**, 213–231.

Andrade, Sandro C., di Pietro, Vadim, & Seasholes, Mark S. 2005. *Understanding the Profitability of Pairs Trading*. UC Berkeley Working Papers. UC Berkeley.

Avellaneda, Marco, & Lee, Jeong-Hyun. 2010. Statistical arbitrage in the US equities market. *Quantitative Finance*, **10**, 1–22.

Barberis, Nicholas. 2000. Investing for the Long Run when Returns Are Predictable. *Journal of Finance*, **55**(1), 225–264.

Broussard, John Paul, & Vaihekoski, Mika. 2012. Profitability of pairs trading strategy in an illiquid market with multiple share classes. *Journal of International Financial Markets, Institutions and Money*, **22**(5), 1188 – 1201.

Caldeira, João F., & Portugal, Marcelo S. 2010. Estratégia Long-Short, Neutra ao Mercado, e Index Tracking Baseadas em Portfólios Cointegrados. *Revista Brasileira de Finanças*, **8**(4), 469–504.

Campbell, John Y., Lo, Andrew W., & MacKinlay, A. Craig. 1997. *The Econometrics of Financial Markets*. Princeton, New Jersey: Princeton University Press.

Carcano, Giovana, Falbo, Paolo, & Stefani, Silvana. 2005. Speculative trading in mean reverting markets. *European Journal of Operational Research*, **163**(1), 132–144.

Chan, Ernest P. 2009. *Quantitative Trading: How to Build Your Own Algorithmic Trading Business*. New Jersey: Wiley trading series.

DeMiguel, Victor, Garlappi, Lorenzo, Nogales, Francisco J., & Uppal, Raman. 2009. A generalized approach to portfolio optimization: improving performance by constraining portfolio norms. *Management Science*, **55**(5), 798–812.

Dunis, Christian L., & Ho, Richard. 2005. Cointegration portfolios of european equities for index tracking and market neutral strategies. *Journal of Asset Management*, **1**, 33–52.

Dunis, Christian L., Giorgioni, Gianluigi, Laws, Jason, & Rudy, Jozef. 2010. *Statistical Arbitrage and High-Frequency Data with an Application to Eurostoxx 50 Equities*. CIBEF Working Papers. CIBEF.

Elliott, Robert J., van der Hoek, John, & Malcolm, William P. 2005. Pairs trading. *Quantitative Finance*, **5**(3), 271–276.

Fung, William, & Hsieh, David A. 1999. Is mean-variance analysis applicable to hedge funds? *Economics Letters*, **62**(1), 53–58.

Gatarek, Lukasz T., Hoogerheide, Lennart F., & Dijk, Herman K. Van. 2011 (Feb.). *A simulation-based Bayes' procedure for robust prediction of pairs trading strategies*. Tinbergen Institute Discussion Papers 09-061/4. Tinbergen Institute.

Gatev, Evan, Goetzmann, William N., & Rouwenhorst, K. Geert. 2006. Pairs Trading: Performance of a Relative Value Arbitrage Rule. *The Review of Financial Studies*, **19**, 797–827.

Goetzmann, Willian N., Ingersoll, Jonathan E., Spiegel, Matthew I., & Welch, Ivo. 2002 (Feb.). *Sharpening Sharpe Ratios*. Yale School of Management Working Papers ysm29. Yale School of Management.

Guidolin, Massimo, Hyde, Stuart, McMillan, David G., & Ono, Sadayuki. 2009. Non-linear predictability in stock and bond returns: When and where is it exploitable? *International Journal of Forecasting*, **25**(2), 373–399.

Hamilton, James D. 1994. *Times Series Analysis*. New Jersey: IE-Princeton.

Hansen, Peter R. 2005. A Test for Superior Predictive Ability. *Journal of Business & Economic Statistics*, **23**(October), 365–380.

Hendry, David F., & Juselius, Katarina. 2001. Explaining cointegration analysis: part II. *Energy Journal*, **22**, 75–120.

Jacobs, Bruce I., & Levy, Kenneth. 1993. Long-Short Equity Investing. *Journal of Portfolio Management*, **1**, 52–64.

Johansen, Soren. 1995. *Likelihood-Based Inference in Cointegrated Vector Autoregressive Models*. USA: Oxford University Press.

Khandani, Amir E., & Lo, Andrew W. 2007. What happened to the quants in August 2007? *Journal of Investment Management*, **5**, 5–54.

Lim, G. C., & Martin, Vance L. 1995. Regression-based cointegration estimators with applications. *Journal of Economic Studies*, **22**(1), 3–22.

Litterman, Bob. 2003. *Modern Investment Management - An Equilibrium Approach*. New York: Wiley.

Lo, Andrew W., & MacKinlay, A. Craig. 1988. Stock Market Prices do not Follow Random Walks: Evidence from a Simple Specification Test. *Review of Financial Studies*, **1**(1), 41–66.

Lo, Andrew W., & MacKinlay, A. Craig. 1990. When Are Contrarian Profits Due to Stock Market Overreaction? *Review of Financial Studies*, **3**(2), 175–205.

Lo, Andrew W., & MacKinlay, A. Craig. 1997. Maximizing Predictability In The Stock And Bond Markets. *Macroeconomic Dynamics*, **1**(01), 102–134.

Lucas, Andre. 1997 (Jan.). *Strategic and tactical asset allocation and the effect of long-run equilibrium relations*. Serie Research Memoranda 0042. VU University Amsterdam, Faculty of Economics, Business Administration and Econometrics.

Nath, Purnendu. 2006 (November). *High Frequency Pairs Trading with U.S Treasury Securities: Risks and Rewards for Hedge Funds*. Working Paper Series. London Business School.

Perlin, Marcelo. 2009. Evaluation of pairs trading strategy at the Brazilian financial market. *Journal of Derivatives & Hedge Funds*, **15**(2), 122–136.

Pole, Andrew. 2007. *Statistical Arbitrage: Algorithmic Trading Insights and Techniques*. USA: Wiley Finance.

Politis, Dimitris N., & Romano, Joseph P. 1994. The stationary bootstrap. *Journal of the American Statistical Association*, **89**(428), 1303–1313.

Poterba, James M., & Summers, Lawrence H. 1988. Mean reversion in stock prices : Evidence and Implications. *Journal of Financial Economics*, **22**(1), 27–59.

Serban, Alina F. 2010. Combining mean reversion and momentum trading strategies in foreign exchange markets. *Journal of Banking & Finance*, **34**(11), 2720–2727.

Soerensen, Jan Tang. 2006. *Hedge Fund Portfolio Construction*. M.Phil. thesis, The Fletcher School - Tufts University.

Stock, James H., & Watson, Mark W. 1988. Testing for Common Trends. *Journal of the American Statistical Association*, **83**, 1097–1107.

Triantafyllopoulos, Kostas, & Montana, Giovanni. 2011. Dynamic modeling of mean-reverting spreads for statistical arbitrage. *Computational Management Science*, **8**(1), 23–49.

Tsay, Ruey S. 2010. *Analysis of Financial Time Series*. New Jersey: John Wiley & Sons.

Vidyamurthy, Ganapathy. 2004. *Pairs Trading, Quantitative Methods and Analysis*. Canada: John Wiley & Sons.

West, Kenneth D. 1996. Asymptotic Inference about Predictive Ability. *Econometrica*, **64**(5), 1067–1084.

White, Halbert. 2000. A Reality Check for Data Snooping. *Econometrica*, **68**(5), 1097–1126.

Zebedee, Allan A., & Kasch-Haroutounian, Maria. 2009. A closer look at co-movements among stock returns. *Journal of Economics and Business*, **61**(4), 279–294.

Appendix: Reality Check Test Statistic

White's Reality Check method uses the test statistic

$$T^{RC} = \max_{k=1,\dots,L} n^{\frac{1}{2}} \overline{f}_k,$$

with

$$\overline{f}_k = n^{-1} \sum_{t=1}^{n} \hat{f}_{k,t+\tau},$$

to test the null hypothesis that

$$H_0 : \max_{k=1,\dots,L} E[f_k] \leq 0.$$

The test statistic is derived from an asymptotic destitution. West (1996) proves that

$$n^{\frac{1}{2}} \left(\overline{f}_k - E[f_k] \right) \xrightarrow{d} N(0, \Omega)$$

where Ω is the $(L \times L)$ asymptotic variance covariance matrix

$$\Omega = \lim_{n \to \infty} var \left\{ n^{\frac{1}{2}} \sum_{t=1}^{n} f_k(Z, \beta_k) \right\}$$

$f_k(Z, \beta_k)$ is the general form of the selection criterion, which measures performance of model k relative to the benchmark, conditional on a given set of data Z. The trick now is that the above expression simplifies to $n^{\frac{1}{2}} \overline{f}_k$ if we impose the element of the null hypothesis where $E[f_k] = 0$ for all k. Remember that the null hypothesis says that all alternative models perform worse or equal to the benchmark model. This means that the element of the null the least favorable to the alternative is that all the models are equally as good as the benchmark, i.e. $\overline{f}_k = 0$. Thus this simplification leads k to a very conservative test statistic.

Continuing on his quest to find a feasible distribution of $n^{\frac{1}{2}} \overline{f}_k$ White states that

$$n^{\frac{1}{2}} \left(\overline{f}_{k,b}^* - \overline{f}_k \right) \xrightarrow{d} n^{\frac{1}{2}} \left(\overline{f}_k - E[f_k] \right)$$

where $\overline{f}_{k,b}^*$ are the bootstrapped values for model k. That this is true can be seen since as n increases

$$\overline{f}_k \xrightarrow{p} E[f_k] \qquad \text{and} \qquad \overline{f}_{k,b}^* \xrightarrow{p} \overline{f}_k$$

for large b. Thus by repeatedly drawing resamples of $\left(\overline{f}^*_{k,b} - \overline{f}_k\right)$ we build up an estimate of the desired distribution $N(0, \Omega)$ of the test statistic in the first equation.

We can now get the Reality Check p-value, taking the model search into account, immediately by comparing the test statistic to the distribution of

$$\max_{k=1,\ldots,L} n^{\frac{1}{2}} \left(\overline{f}^*_{k,b} - \overline{f}_k\right),$$

to find the according percentile.