UNIVERSITA' DEGLI STUDI DI PADOVA

DIPARTIMENTO DI SCIENZE ECONOMICHE ED AZIENDALI "M. FANNO"

DIPARTIMENTO DI MATEMATICA "TULLIO LEVI-CIVITA"

CORSO DI LAUREA MAGISTRALE IN ECONOMICS AND FINANCE

TESI DI LAUREA

"PAIRS TRADING STRATEGIES: A COINTEGRATION-BASED APPROACH"

RELATORE:

CH.MO PROF. CLAUDIO FONTANA

LAUREANDO: EDOARDO PIALLINI

MATRICOLA N. 1179473

ANNO ACCADEMICO 2019 – 2020

Il candidato dichiara che il presente lavoro è originale e non è già stato sottoposto, in tutto o in parte, per il conseguimento di un titolo accademico in altre Università italiane o straniere.
Il candidato dichiara altresì che tutti i materiali utilizzati durante la preparazione dell'elaborato sono stati indicati nel testo e nella sezione "Riferimenti bibliografici" e che le eventuali citazioni testuali sono individuabili attraverso l'esplicito richiamo alla pubblicazione originale.

The candidate declares that the present work is original and has not already been submitted, totally or in part, for the purposes of attaining an academic degree in other Italian or foreign universities. The candidate also declares that all the materials used during the preparation of the thesis have been explicitly indicated in the text and in the section "Bibliographical references" and that any textual citations can be identified through an explicit reference to the original publication.

Firma dello studente

____

Contents

List of Figures

List of Tables

Introduction

In this dissertation we examine a popular quantitative investment strategy commonly referred to as statistical pairs trading, which was first pioneered in the mid 1980's by a quantitative trading group headed by Nunzio Tartaglia at Morgan Stanley (Gatev *et al.*, 2006).

Pairs trading works by exploiting profitable opportunities arising from temporary mispricing between prices of related securities which share a long-term equilibrium relationship. In the presence of mispricing, one security will be overvalued relative to the other security. Pairs trading strategies aim at exploiting this mispricing by simultaneously selling the overvalued security (short position) and purchasing the undervalued security (long position). The trade is then closed out by taking opposite positions on these securities, i.e. by selling the long position and off-setting the short position, when the security prices have settled back to their long-term equilibrium, leading to a profit for the trader (Huck and Afawubo, 2015).

Pairs trading is classified within a group of quantitative trading approaches collectively referred to as statistical arbitrage strategies. The arbitrage part in this context is somewhat misleading as arbitrage implies a riskless profit opportunity at zero upfront net investment. Clearly a pairs trading strategy is by no means riskless since there is not guarantee that the securities' prices will converge to the long-term equilibrium value (Avellaneda and Lee, 2010). Lazzarino *et al.* (2018) define statistical arbitrage as a relative value strategy with positive expected return and a tolerably small potential loss, highlighting the fact that the risk of potential losses is a crucial component of this type of strategies, which differentiate them from pure arbitrage strategies where negative payoffs are assumed to occur with zero probability.

In this thesis we focus on cointegration-based pairs trading strategies in which the pairs selection process is based on the concept of cointegrated prices which possess a stationary long-term equilibrium relationship with the associated property of mean reversion. The phenomenon of cointegration has been observed for the first time by Engle and Granger in 1987. The idea is very simple, even if two time series are found to be nonstationary, in some instances it is possible that a linear combination of them is stationary. If this is the case the two time series are said to be cointegrated. In other words, this means that it is possible that two time series, which are found to be individually nonstationary, tend to move together over time, in the sense that they cannot drift too far away from each other, except for transitory fluctuations. This cointegration relationship defines a sort of long-term equilibrium relationship to which the time series will be forced to return, despite short run deviations (Vidyamurthy, 2004). The aim of this thesis is to explore the profitability of different cointegration-based pairs trading strategies using the daily closing stock prices of the major banks in the Italian banking system over the period from 2 January 2015 to 30 December 2019. Our findings indicate that almost all the strategies that we have analysed result to be significantly profitable under all the specifications analysed. Moreover, we investigate if the impact of losses, due to the closure of the positions at the end of the trading period before the convergence to the closing trigger has occurred, could be reduced if a longer trading period is taken into account. Our results show that an increase in the length of the trading period significantly reduces the number of unprofitable trades, leading to greater average annualized returns for all the parametrization examined.

The dissertation is organized as follows. The first chapter is devoted to the introduction of the various classes of statistical arbitrage strategies, with particular emphasis on pairs trading. Specifically, we analyse the different pairs trading approaches cited in the literature, which are: distance approach, cointegration approach, time series approach, stochastic control approach and copula approach. The second chapter is an accurate presentation of all the statistical and econometric elements needed to understand the concept of cointegration, as exposed by Engle and Granger in 1987. The third chapter is dedicated to the analysis of the three steps required for the implementation of a cointegration-based pairs trading strategy, which are pre-selection of stock pairs, testing for cointegration and trading design. The fourth chapter is dedicated to our empirical analysis and the simulations, along with the main results of this research.

Chapter 1

Statistical Arbitrage Pairs Trading Strategies: Overview

1.1 Introduction

According to Ross (2004, p. 1) the arbitrage is an investment strategy that guarantees positive payoffs in some circumstances with no possibility of negative payoffs and without initial net investment. A widely applied arbitrage tactic entails the sale of a security at a relatively high price in one market and the simultaneous purchase of the same security, or its functional equivalent, in another market at a relatively low price (Sharpe *et al.*, 1999, p. 284).

The principle is very simple, the arbitrageur extracts a riskless profit exploiting the temporary discrepancies in the price of a security in different markets; he purchases the security in the market in which the price is lower and simultaneously he sells it in the market in which the price is higher.

The role of arbitrage is crucial in the analysis of securities markets, its effect is dual: it allows to bring the securities' prices to their fundamental values and to keep markets efficient (Shleifer and Vishny, 1997).

The assumption of no arbitrage represents a fundamental pillar of financial economics and mathematical finance. Indeed, as argued by Ross (2004, p. 2), no arbitrage is a necessary condition for an equilibrium in the financial markets: if arbitrage opportunities were present in the market, then the demand and supply for the securities involved would be infinite, which is inconsistent with equilibrium.

Analytically the absence of arbitrage condition can be described as follow:

If $\mathbb{P}(\bar{r} \neq r_f) > 0$ then it must be true that $\mathbb{P}(\bar{r} > r_f) > 0$ and $\mathbb{P}(\bar{r} < r_f) > 0$

Where $\bar{r}$ is the return of the portfolio composed of available assets and r_f is the return of the risk-free asset.

The meaning of this condition is that none of the two assets systematically produces a higher return than the other, thus it is not possible to gain a profit without the risk of incurring in a loss.

According to Shleifer and Vishny (1997) pure arbitrage, thus an investment strategy which requires no capital and entails no long run fundamental risk and which always leads to a positive payoff, is unlikely to occur in real trading environment, since in most real world situations arbitrageurs in order to arrange their trades need substantial amounts of capital and face some forms of risk. In other terms, their positions pay off only on average and not with certainty. When the risks, faced by the trader in arranging an investment strategy, are statistically assessed

it is appropriate to use the term statistical arbitrage.
Statistical arbitrage is not an investment strategy free of risk (i.e. it is not a pure arbitrage strategy), but it is an investment strategy in which the risks is assessed using statistical tools. Lazzarino *et al.* (2018) define the statistical arbitrage as a relative value strategy with positive expected excess return and an acceptably small potential loss. This conceptual definition underlines different important features of statistical arbitrage: it is defined as a relative value strategy, meaning that it is a strategy aimed to find mispricing using historical relationship between securities. Another relevant element is given by the expected positive excess return which highlights two important characteristics of statistical arbitrage: first, the fact that losses are allowed, since the focus is on expected return and not on positive payoff; this is crucial in order to differentiate statistical arbitrage from pure arbitrage where negative outcomes are not admissible. Second, the fact that arbitrageurs invest in a strategy involving some risk only if there are expectations of return higher than the risk free, i.e. positive expected excess return, whenever an initial investment is required. The last element is the acceptability of small potential loss; according to Lazzarino *et al.* (2018) this feature is crucial to differentiate statistical arbitrage from simple investment. A strategy in order to be classified as arbitrage needs to have a constrained loss profile, meaning that the strategy is closed whenever the prearranged criteria are no longer satisfied, e.g. when the loss is no longer acceptably small or when the expected excess return is no longer positive.
In other words, statistical arbitrage is an investment strategy in which one can accept negative pay-outs with a small probability if the expected positive returns are high enough and the probability of losses, assessed using statistical techniques, is sufficiently small (Saks and Maringer, 2008).
Lazzarino *et al.* (2018) proposed the following classification of statistical arbitrage strategies in fixed income, based on the previous categorization introduced by Duarte *et al.* (2006):

- *Yield curve arbitrage* (or *Term structure arbitrage*)
- *Volatility arbitrage*
- *Mortgage arbitrage*
- *Swap spread arbitrage*
- *Capital structure arbitrage* (or *Credit arbitrage*)
- *Pairs trading*

The *yield curve arbitrage* is a typical statistical arbitrage strategy which involves taking market-neutral long-short positions at different point along the yield curve, as suggested by a relative value analysis (Lazzarino *et al.*, 2018). The yield curve is a graphical representation of the yields of fixed income treasury securities having equal credit quality but different maturities. When the yield curve is flat this means that short-term yields and long-term yields are similar, while when the yield curve is heavily sloped there is a significant gap between short-term and long-term yields. The yield curve arbitrage seeks to gain from treasuries misprices along different points of the yield curve, which represent profit opportunities for the investor, by taking long and short positions in bonds with different maturities in such a way that the risk of the portfolio is minimized.

The *volatility arbitrage* is a widely used trading strategy that seeks to profit from the difference between actual volatility, that is the amount of 'noise' in the stock market or the amount of randomness that transpires, and the implied volatility, i.e. how the market is currently pricing the option based on the stock (Ahmad and Willmott, 2005). Since the option pricing is influenced by the volatility of the underlying stock, if the actual volatility and the implied volatility differ, there will be a discrepancy between the expected price of the option and its market price which could generate profitable opportunities. According to Duarte *et al.* (2006) the simplest form of volatility arbitrage can be implemented through a delta-neutral portfolio obtained selling an option (short position in the option) and then delta-hedging the exposure to the underlying asset (long or short position in the underlying asset). The investor, in this situation, hopes to profit from the tendency of implied volatility to exceed subsequent realized volatility. Assuming that the price of the stock does not change, if the investor is correct about implied volatility declining, he/she may profit from the reduction in the value of the option. *Mortgage backed-securities arbitrage* is a strategy that consists of buying mortgage backed-securities (MBSs) while hedging their interest rate exposure primarily through derivatives. A mortgage-backed security is a securitization of a set of mortgages collateralized by real estate. Through the securitization the set of mortgages held by a financial institution is pooled and sold to investors which become the recipients of the cash flows generated by the mortgages (principal and interest payments). From the investor's perspective, the MBS is a fixed-income security embedding a prepayment option: homeowners can choose to prepay all or part of their loans at any time during the life of the mortgages, making the mortgage's future cash flow, and therefore the MBS value, uncertain. Specialized investors using proprietary models can estimate the option-adjusted spread (OAS) that is the security's incremental value with respect to Treasury bonds with the same maturity, adjusted for impact of possible MBS prepayments. At this point, MBSs offering the highest OAS values, that are the cheapest MBSs, are purchased and hedged with short sale of Treasury bonds of equal duration or with the sale of Treasury bond futures, establishing a position hedged against the interest rate risk (Stefanini, 2006, pp. 173-174).

A *swap spread trade* is a strategy in which the investor bets on the difference between a fixed and a floating yield. Following Duarte *et al.* (2006), it is structured in two parts: on the one hand, the trader enters into a par swap and receives a fixed coupon rate *CMS* (constant maturity swap rate) and pays the floating Libor rate (L_t). On the other hand, the trader shorts a par Treasury bond (*CMT*, i.e. constant maturity Treasury rate) with the same maturity of the swap and invests the revenues in an account earning the repurchase agreement rate (r_t). Combining the two parts shows that the trader receives (*CMS -CMT*) that is the fixed interest rate component and pays (L_t - r_t) which represents the floating spread. The swap spread arbitrage is a bet on whether the fixed component received by the investor will be larger than the floating spread paid.

Capital structure arbitrage is a trading strategy that seeks to take advantage of the relative mispricing between a company's debt and its other securities, such as equity. The rationale for this strategy is to exploit lack of integration or synchronicity between different securities issued by the same company. According to Lazzarino *et al.* (2018) a simple version of capital structure arbitrage can be implemented exploiting the mispricing between a company's credit default swap[1] (CDS) and its equity. Using information about the equity price and the capital structure of an obligor, the investor computes the theoretical CDS spread that is then compared with the level quoted in the market. A discrepancy between the two values is the signal of a potential profitable opportunity for the investor. If the market spread is lower (higher) than the theoretical spread, then the profitable strategy will be a long (short) position on the CDS contract while simultaneously hedging the equity with a long (short) position.

The simplest statistical arbitrage strategy is the so-called statistical *pairs trading*. The idea behind this strategy is trivial: find two securities whose price is affected by the same common factors. According to the law of one price securities with similar characteristics should have similar prices. If for some reasons the price of the securities diverges: sell the higher-priced and buy the lower-priced, in the expectation that the mispricing will correct itself in the future. One of the main objectives of a trader investing in the financial markets is to gain a profit and to do so he must be able to sell overvalued securities and buy undervalued ones. The problem is that, in order to understand if a security is undervalued or overvalued, the trader must know

[1] Credit default swaps are essentially insurance contracts against the default of an obligor. In particular, the buyer of the CDS contract pays a premium to the seller, usually express as a percentage of the notional amount of the underlying bond, and the seller agrees to pay the notional value of the bond should the obligor default before the maturity of the contract (Duarte *et al.*, 2006).

the fundamental value of that security and this is a very difficult, or even impossible, task. With the introduction of pairs trading strategy, the focus moves from absolute pricing (fundamental value) to relative pricing. When dealing with pairs trading is not important that prices are correct, the only thing that matters is that similar securities have similar prices (Vidyamurthy, 2004, p. 74).

The principal assumption of pairs trading is the existence of a long-term equilibrium relationship between two securities, since pairs trading is a short-term speculation strategy, modelling this relationship would allow the trader to take advantage of any short-term deviations opening a long-short position that will be reversed upon restoration of the price relationship, thus leading to a profit (Rad *et al.*, 2016).

Put differently, pairs trading exploits opportunities generated by temporary anomalies between prices of related assets which have a long-term equilibrium. When such an event occurs, one asset will be overvalued compared to the other one. At this point, the trader will invest in a two-assets portfolio composed by a short position in the overvalued asset and a long position in the undervalued asset, it follows that pairs trading can be classified as a contrarian investment strategy since the investor is buying a stock performing relatively poorly and he/she is selling a stock performing relatively well. The trade is closed out by reverting the positions after the asset prices have settled back into their long-term relationship (Puspaningrum, 2012).

From what has been said, it follows that pairs trading is a market-neutral investment strategy since the return from the strategy is uncorrelated with the movement of the market, i.e. the trader profits from the short-term discrepancies in the prices of the assets considered regardless of whether the market goes up or down (Vidyamurthy, 2004, p. 8).

However, the implementation of this strategy is far more complicated, since there are a lot of factors that need to be considered, such as the risk-attitude of the traders and their capability of bearing losses. The problem of pairs trading strategy is that the existence of a long-term relationship between two securities does not say anything about the timing of mean-reversion (see Section 2.10.1) or about the width of the divergence in price. What can be said is that it is reasonable to assume that given enough time the relationship between two similar securities will be restored, but there is no chance to know *a priori* the patterns that they will follow, and so it is possible that the traders may suffer considerable losses. According to Krauss (2017) pairs trading research is divided in five streams of literature:

- *Distance approach*: searching for pairs by minimizing the sum of squared differences between two normalized time series over a fixed formation period and define opening and closing threshold as trading signals (Gatev *et al.*, 2006).

- *Cointegration approach:* pairs selection is based on the concept of cointegrated price series which possess a stationary long-term equilibrium relationship with the associated property of mean reversion. Whenever a deviation from the long-run mean arise, a profitable opportunity is created (Lin *et al.,* 2006).
- *Time-series approach:* this approach explicitly models the mean reverting behaviour of the spread, i.e. the difference between two stock prices, in a continuous time setting. It relies on the assumption that the spread follows an Ornstein-Uhlenbeck process, which is useful for the generation of optimized trading signals using different methods of time-series analysis.
- *Stochastic control approach:* the focus lies on finding the optimal investment in the two legs of a pair when other assets are available. Stochastic control theory is used to obtain closed form characterisations of the portfolio optimization problem (Krauss, 2017).
- *Copula approach:* this approach has been introduced in order to overcome the limitation of correlation or cointegration as a measure of dependency. Copulas are useful extensions of approaches for modelling joint distributions and dependence between financial assets (Liew and Wu, 2013).

1.2. Distance Approach

The most important research concerning the distance approach have been conducted by Gatev *et al.* (2006) and Do and Faff (2010). The pairs trading strategy developed by Gatev *et al.* (2006), and subsequently used by Do and Faff (2010), followed a two steps process:

1. Creation of pairs over a 12-month formation period.
2. Trade the pairs in the next 6-month trading period.

The first step can be further broken down into two phases: first, a cumulative total return index is constructed for each stock and normalized to the first-day of the formation period; second, considering a portfolio of *n* different stocks, in the CRSP (Centre for Research in Security Prices) universe, the sum of Euclidean squared distances (SSD) for the price time series of *n(n-1)/2* possible combinations of pairs is calculated and then each stock is paired with the 'matching partner' that minimizes the SSD (Krauss, 2017):

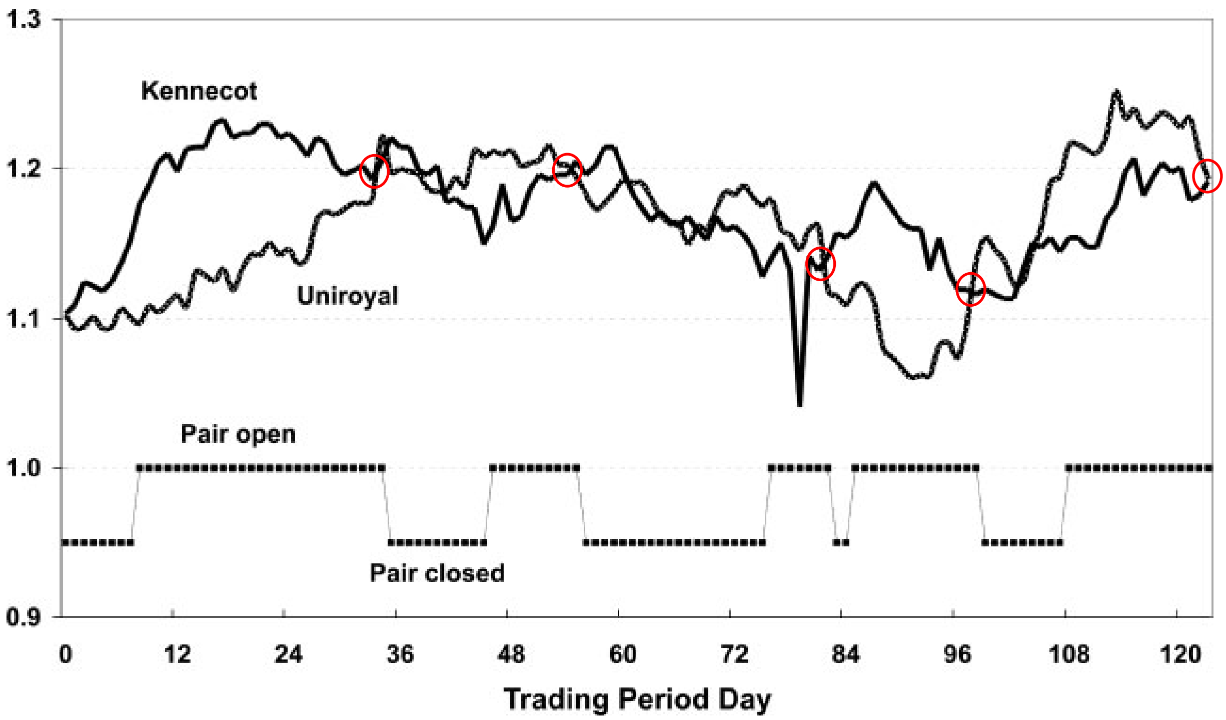
$$SSD_{i,j} = \sum_{t=1}^{T} (P_{i,t} - P_{j,t})^2$$

with $P_{i,t}$ and $P_{j,t}$ the normalized prices for stock *i* and *j* on day *t*, and *T* the number of trading days in the formation period.

For the trading step only the 20 pairs with the smallest historical distance measure are considered. The trading rule for opening positions is based on a standard deviation metric: positions are opened when prices diverge by more than two historical standard deviations, as estimated during the pairs formation period, and they are closed at the next crossing of the prices (i.e. when the spread between the two prices equals zero), at the end of the trading interval or upon delisting (Gatev *et al.*, 2006).

Figure 1.1 illustrates the pairs trading strategy adopted by Gatev *et al.* (2006) using two stocks (Uniroyal and Kennecott), in the six-month period from August 1962 to January 1963. The top lines represent the normalized price path with dividend reinvested, while the bottom line indicates the opening and closing of the strategy on a daily basis. Notice that the position first opens around the seventh trading day, when the spread between the prices is sufficiently large, and it closes on day 36, when the prices converge. While over the first interval, the spread initially increased significantly before convergence, in the subsequent intervals the prices remain relatively close and cross frequently.

Figure 1.1: *Daily Normalized Prices: Kennecott and Uniroyal (Trading period August 1963-January 1964)*



Source: *Gatev et al. (2006)*

The total payoff of the trader is the algebraic sum of the cash flows occurred during the trading period. During the trading period pairs are opened when the spread between the prices of the stocks considered diverges by more than two standard deviations, and then positions can be unwound for three reason:

- *Convergence*: pairs that open and converge during the same trading interval will generate positive cash flows. Since pairs can reopen after initial convergence, they can have multiple positive cash flows (multiple round-trip trades) during the same trading period.
- *Last day of the trading period:* pairs that open but never converge will only have cash flows on the last day of the trading period when all positions are closed out. The cash flows generated by this type of trades can be either positive or negative depending on the spread between prices on the last day of the trading period.
- *Delisting:* if a stock in a pair is delisted from CRSP, the position in the pair involving the stock is closed using the last available price. Even in this case, the cash flows can be either positive or negative depending on the spread between the prices at the date of the delisting.

The results obtained by Gatev *et al.* (2006) are remarkable: in the period from January 1962 to December 2002 the excess monthly return for the 20 best pairs is 1.436% (t-Statistic= 11.56), that is both economically and statistically significant. As argued by the authors, pairs trading is a contrarian investment strategy, thus the returns may be biased upward because of the bid-ask bounce. Bid-ask bounce occurs when the price of a stock bounces back and forth within the very limited range between the bid price and the ask price. According to Alexander (2008, p. 331) this happens when market are not trending and there is roughly the same proportion of buyers and sellers in the market so that the traded price tends to flip between the bid and the ask prices, without implying a real movement in the price of the stock.

Pairs trading tends to buy stocks whose prices have been decreasing, thus closing transactions of many of these stocks may be at bid prices due to substantial selling pressure. It also tends to sell stocks whose prices have been increasing, thus the closing transactions of many of these stocks may be at ask prices due to significant buying pressure. Thus, it is possible that some portion of the trading profits is due to the bid-ask bounce, rather than to actual returns associated with pairs trading (Mori and Ziobrowski, 2011).

In order to minimize the effect of the bid-ask bounce the excess monthly returns are re-calculated considering the situation in which positions are opened on the day following the divergence of the spread by more than two historical standard deviations, and they are liquidated on the day after the convergence of the prices. The drop in the monthly returns is substantial (54.1 basis point) from 1.436% to 0.895% (t-Statistic= 9.29), suggesting that an important portion of the excess returns may be due to the bid-ask bounce. Nonetheless the results remain economically and statistically significant.

Gatev *et al.* (2006) also demonstrate that pairs trading is profitable in every sector by restricting the matching rules in order to obtain pairs composed by stocks belonging to the same large sector grouping used by S&P (Utilities, Transportation, Financial and Industrials), the monthly excess returns remain significant, ranging from 1.084% (t-Statistic=10.26) in the Utilities sector to 0.577% (t-Statistic= 4.26) in the Transportation sector.

Despite the positive and significant excess returns, what emerges from a subperiod analysis is that the profitability of pairs trading is declining over time. According to Gatev *et al.* (2006) the monthly excess return of the top 20 pairs drops from 1.181% in the period between 1962 and 1988 to 0.375% in the period between 1989 and 2002, representing a 68% decline. These results are in line with those by Do and Faff (2010) that, considering the same time interval, registered a reduction of the monthly excess return from 0.86% to 0.37% (57 percent decline).

The purpose of the study conducted by Do and Faff (2010) is to find a possible explanation for the decline in the returns, since the answer provided by Gatev *et al* (2006) concerning a latent risk factor, that is not captured by conventional measures of systemic risk, remained relatively dormant in the second part of the sample (January 1989-December 2002), is not considered exhaustive. The results by Do and Faff (2010) show that the main driver of the general declining trend in pairs trading profitability is the increasing proportion of nonconvergent pairs (pairs that do open but never converge during the trading interval) from 26% (1962-1988) to 39% (1989-2002) and to 40% (2003-2009) at the expense of the profitable pairs with multiple round-trip trades (pairs that open several times within the same trading interval) that decreased from 42% (1962-1988) to 24% (1989-2009). Interestingly, Do and Faff (2010) found out that pairs trading performance is particularly strong during market crises: dot-com bubble (2000-2002) and financial crisis (2007-2009). During the dot-com bubble the good performance of pairs trading strategy was driven by higher average monthly returns among convergent pairs with respect to adjacent periods: 4.84% in the crisis period compared with 1.51% of the period 1989-1999 and 1.69% of the period 2003-2007. On the other end, the profitability of pairs trading during the recent financial crisis was driven by a combined effect of higher average monthly return from 1.69% in the pre-crisis interval 2003-2007 to 2.49% in the crisis period 2007-2009 and also to a decrease in the proportion of divergent pairs from 44% (2003-2007) to 32% (2007-2009) accompanied by an increase in the fraction of multiple-round trip pairs from 18% (2003-2007) to 37% (2007-2009) (Do and Faff, 2010).

The distance approach developed by Gatev *et al.* (2006) has several advantages: it is easy to implement, robust to data snooping[2] and results in statistically significant risk-adjusted excess returns (Krauss, 2017). Moreover, as Do *et al.* (2006) point out the distance approach is economic model-free and consequently, it has the advantage of not being exposed to model mis-specification and mis-estimation. However, being a non-parametric approach, it lacks forecasting ability regarding the convergence time or the expected holding period.

As argued by Galenko *et al.* (2012) another shortfall in the study by Gatev *et al.* (2006) is that they support their results using the theory of cointegration asset prices (the idea is that cointegrated systems have a long-run equilibrium, thus two time series that are cointegrated are also mean-reverting (see section 2.10.1)), but never used any tests for cointegration to justify the trading strategy. The omission of the cointegration test implies that there is not a rational reason for expecting a long-term equilibrium; in other words, there is no motive to believe that prices that have diverged will converge again. Clearly the potential lack of a long-term equilibrium implies higher divergence risks which in turns leads to higher potential losses. Finally, according to Krauss (2017), the choice of Euclidean squared distance as selection metric is analytically suboptimal. A rational trader has the objective of maximizing the excess return per pair, which is the product between the number of trades per pair and the profit per trade. The profit-maximizing investor seeks for spread exhibiting frequent and significant divergences (high spread variance) from the equilibrium with subsequent convergences (strong mean-reversion) to it. The empirical spread variance and the average sum of squared distances can be expressed as follow:

$$s^2_{P_i-P_j} = \frac{1}{T}\sum_{t=1}^{T}(p_{i,t}-p_{j,t})^2 - \left(\frac{1}{T}\sum_{t=1}^{T}(p_{i,t}-p_{j,t})\right)^2$$

$$\overline{SSD}_{P_i,P_j} = \frac{1}{T}\sum_{t=1}^{T}(p_{i,t}-p_{j,t})^2 = s^2_{P_i-P_j} + \left(\frac{1}{T}\sum_{t=1}^{T}(p_{i,t}-p_{j,t})\right)^2$$

[2] Data snooping is a statistical bias that occurs when a given set of data is used more than once for purpose of inference. When such data reuse occurs, there is the possibility that any satisfactory results may be due to chance rather than to any merit inherent to the method used. The probability that a result arises by chance increases with the number of combinations tested (White, 2000).

with $p_{i,t}$ and $p_{j,t}$ denoting the realization of the normalized price processes $P_i = (P_{i,t})_{t\epsilon T}$ and $P_j = (P_{j,t})_{t\epsilon T}$ and $s^2(.)$ the sample variance.

Recalling that Gatev *et al.* (2006) choose the top 20 pairs with minimum SSD, it is easy to see that the optimal pair, that is the one with zero SSD, has a spread of zero and thus produces no profits. This is a clear signal that the choice of Euclidean squared distance as selection metric is not optimal, since one would expect the best pair to generate the highest profits.

1.3. Cointegration approach

The phenomenon of cointegration has been observed for the first time by Engle and Granger (1987). Even though two time series are nonstationary it is possible that in some instances, a linear combination of the two is stationary, in this case the time series are said to be cointegrated (Vidyamurthy, 2004, p. 76).

Consider $\{x_t\}$ and $\{y_t\}$ to be time series integrated of order one, i.e. $x_t{\sim}I(1)$ and $y_t{\sim}I(1)$ (see Section 2.5). If there is a linear combination of the two time series that is stationary (see Section 2.2), i.e. $(x_t - \beta y_t){\sim}I(0)$, it is possible to conclude that $\{x_t\}$ and $\{y_t\}$ are cointegrated (see Section 2.10.1).

Alexander (2008, p. 201) argued that there is little incentive in building forecasting models based on nonstationary processes since they are totally unpredictable over time (see Section 2.6); conversely stationary processes can be used for this purpose, by exploiting their mean-reverting behaviour. Cointegration is a measure of long-term dependency between asset prices: whenever a spread is found to be mean-reverting this means that the two asset prices are "tied together" in the long-term (see Section 2.10). In other words, while it is not possible to predict exactly where the price of an asset will be in the future, given its nonstationary nature, if two time series are found to be cointegrated, this implies the existence of a long term equilibrium relationship that prevents them from drifting too far apart from each other. In the short-run the spread could deviate from the long-term equilibrium (that is the long-run mean of the linear combination of the two time series), allowing for profitable trading opportunities, but given a sufficiently long timeframe the prices will adjust themselves restoring the equilibrium.

The cointegration relationship can be equivalently shown using the Error Correction Model (see Section 2.10.4). The error correction representation for two nonstationary time-series $\{x_t\}$ and $\{y_t\}$ is:

$$y_t - y_{t-1} = \alpha_y(y_{t-1} - \beta x_{t-1}) + \varepsilon_{y_t}$$

$$x_t - x_{t-1} = \alpha_x(y_{t-1} - \beta x_{t-1}) + \varepsilon_{x_t}$$

where ε_{x_t} and ε_{y_t} are white noise processes associated respectively to $\{x_t\}$ and $\{y_t\}$. Consider the first equation, the left-hand side represents the increment of the time series at each time step. The right-hand side is the sum of the error correction part, i.e. $\alpha_y(y_{t-1} - \beta x_{t-1})$, and the white noise component, i.e. ε_{y_t}. The equation shows that the evolution of a time series consists of a white noise process which is responsible for possible deviations from the long-run equilibrium and an error correction term which reverts the time series towards its long-run equilibrium (Rad *et al.*, 2016).

Pairs trading strategy involves trading on the oscillations about the equilibrium value for the spread: when the spread has diverged "sufficiently" from the equilibrium value, an appropriate position in the two stocks is opened, betting that the divergence will correct itself restoring the equilibrium. The key steps involved in the design and analysis of the pairs trading strategy implemented using the cointegration approach are (Vidyamurthy, 2004, pp. 83-84):

- Identification of stock pairs potentially cointegrated based on statistical analysis of historical data (see Section 3.2).
- Cointegration testing to verify the hypothesis that the stock pairs are indeed cointegrated based on statistical evidence (see Section 3.3).
- Identification of optimal entry/exit threshold using parametric or non-parametric method in order to maximize the expected profits (see Section 3.4).

According to Krauss (2017), the key benefit of the cointegration approach is the econometrically more reliable equilibrium relationship of identified pairs, compared to the distance approach proposed by Gatev *et al.* (2006). Moreover, it provides a statistical methodology for modelling both the long-term and the short-term dynamics (see Section 2.10 and Section 2.11).

1.4. Time-Series Approach

Elliott *et al.* (2005) provide the most cited work for time-series based pairs trading. They model the mean reversion behaviour of the spread between the paired stocks in a continuous time setting. The observed spread, $\{y_k\}$, is defined as the difference between the prices of the paired stocks, and it is driven by a state process $\{x_k\}$ plus a measurement error captured by a Gaussian noise (ω_k):

$$y_k = x_k + \mathrm{D}\omega_k$$

Where x_k represents the value of some real variable at time $t_k = k\tau$, ω_k~IID $N(0,1)$ and D > 0 is a parameter representing the standard deviation of the error term ω_k.

The state process $\{x_k \mid k=0,1,2\ldots\}$ is assumed to be mean reverting:

$$x_{k+1} - x_k = (a - bx_k)\tau + \sigma\sqrt{\tau}\,\varepsilon_{k+1} \qquad (1)$$

with a, $\sigma \in R_0^+$, $b \in R^+$ and $\{\varepsilon_k\}$ consists of *i.i.d.* random variables with standard normal distribution, i.e. N (0, 1). The purpose is to compute the conditional expectation:

$$\hat{x}_k = \boldsymbol{E}\,[x_k|Y_k]$$

where $Y_k = \sigma\{y_0, y_1, y_2, \ldots, y_k\}$ represents the information obtained by observing $y_0, y_1, y_2, \ldots, y_k$. The conditional expectation in the previous formula denotes the best estimate of the hidden state process x at time k, which is the variable of interest, given the information obtained from the observed spread process up to time k.

According to Triantafyllopoulos and Montana (2011), the observed process should be seen as a noisy realization of the underlying hidden process $\{x_k\}$ describing the true spread, which captures the true market condition. Thus, a comparison between the estimated unobserved spread process and the observed one, may lead to the discovery of temporary market inefficiencies. At time t, if $y_t > (<)\ \hat{x}_{t|t-1} = \boldsymbol{E}\,[x_t|Y_{t-1}]$, that is if the observed spread is strictly larger (lower) than the best estimate of the process x_t given the information up to t-1. In this situation, the spread is regarded as too large, and so the investor could take a long (short) in the spread portfolio whenever y_k exceeds $\hat{x}_{k|k-1}$ by some threshold value, calculated using time-series analysis, and profiting when a correction, i.e. when the spread has reverted back to its mean, occurs.

The process $\{x_k\}$ mean reverts to $\mu = {}^{a}/_{b}$ with strength b. Equation (1) can be rewritten as:

$$x_{k+1} = \mathrm{A} + \mathrm{B}x_k + \mathrm{C}\varepsilon_{k+1} \qquad (2)$$

where $\mathrm{A} = \mathrm{a}\tau$, $B = 1 - \mathrm{b}\tau$ and $\mathrm{C} = \sigma\sqrt{\tau}$.

The discrete process described by equation (2) can be approximated by a continuous process $\{X_t \mid t \geq 0\}$ that satisfies the following stochastic differential equation:

$$\mathrm{d}X_t = \rho(\mu - X_t)\mathrm{dt} + \sigma dW_t \qquad (3)$$

with $\mu = {}^{a}/_{b}$ denoting the mean, $\rho = b$ denoting the speed of mean-reversion and $\{W(t) \mid t \geq 0\}$ representing a standard Brownian motion.

Using the Ornstein-Uhlenbeck, described in equation (3), it is possible to compute the most likely time T at which $X(T)=\mu$, representing the moment in which the position is closed (Elliot *et al.*, 2005):

$$T = \frac{1}{\rho}\hat{t} = \frac{1}{2\rho} ln\left[1 + \frac{1}{2}\left(\sqrt{(c^2-3)^2+4c^2} + c^2 - 3\right)\right]$$

In other terms, T is the expected time needed for the process X_t to converge to its long-term equilibrium, assuming that at time t=0 the process diverged from its equilibrium value of a certain prespecified threshold.

Elliott *et al.* (2005) propose a strategy in which the investor enters a pairs trade when the spread moves away from its mean (μ) and hits one of the two bounds, that is when:

$$y_k \geq \mu + c\left(\sigma/\sqrt{2\rho}\right) \qquad \text{or} \qquad y_k \leq \mu - c\left(\sigma/\sqrt{2\rho}\right)$$

where c is a fixed parameter that can be regarded as an optimal threshold to open a pair trade for which Elliot *et al.* (2005) give no indication on how to determine it.

The investor should enter in a pairs trade as long as the spread moves away from its mean, hitting one of the two bounds, knowing that the spread will revert back to its mean since it follows a mean-reverting process. In particular, the position is unwound at time T, which represents the first passage time resulting from the OU process (Krauss,2017). According to Do *et al.* (2006) the model proposed by Elliott *et al.* (2005) has three main advantages:

- It captures mean reversion which underlies pairs trading.
- It can be exploited for forecasting purposes. It is indeed possible to compute the expected time for the spread to converge to the long-term mean.
- The model is fully tractable, meaning that all the parameters can be estimated by the Kalman filter in a state space setting.

Despite the great advantages, Do *et al.* (2006) criticised that the model has a fundamental limitation since it is only applicable to securities in return parity, i.e. in the long run, the two stocks must provide the same return such that any departure from it will be expected to be corrected in the future. This is a severe limitation since in practice it is almost impossible to find two stocks with identical return.

Do *et al.* (2006) proposed a pairs trading strategy, known as *Stochastic Residual Spread*, that models mispricing at return level rather than price level.

This approach differs from the others due to the methodology of quantification of mean reversion behaviour, that is made considering the theoretical asset pricing relationship instead of being purely based on statistical consideration leading to ad hoc trading rules (Do *et al.*, 2006).

Their model starts with an assumption on the existence of a long-term equilibrium in the relative value of the two stocks measured by the spread. Mispricing is described as the state of disequilibrium, which is quantified by a *residual spread function* $G(R_t^A, R_t^B, U_t)$, where R_t^A and R_t^B represents respectively the return of stock A and stock B at time *t*, while *U* denotes some exogenous vector potentially present in formulating the equilibrium.

Do *et al.* (2006) adopt the same modelling framework proposed by Elliot *et al.* (2005), that is a one-factor stochastic model used to describe the state of mispricing or disequilibrium and incorporates a white Gaussian noise ($\omega_t \sim$ IID *N (0,1)*) that affects its actual observation being measured by the residual spread function $G(R_t^A, R_t^B, U_t)$ (Fiz, 2014). The state of mispricing or residual spread with respect to a given long-term equilibrium relationship is described by the variable x_t whose dynamics is follows a Vasicek process:

$$dx_t = \kappa(\theta - x_t)dt + \sigma dB_t$$

where dB_t is a standard Brownian motion, θ is the long-run mean of the state variable x_t and κ is the speed of mean reversion.

Instead, the observed mispricing (residual spread function) is defined as follow:

$$y_t = G_t = x_t + \omega_t$$

These two equations described a state space model of relative mispricing, defined with respect to some equilibrium relationship between two assets.

At this point, Do *et al.* (2006) specify the residual spread function (*G*) as follow (see Do *et al.* (2006)):

$$G_t = G(R_t^A, R_t^B, U_t) = R_t^A - R_t^B - \Gamma r_t^m$$

where Γ is a vector of exposure differentials and r_t^m is the vector of risk factor returns in excess over the risk free asset return.

If the value of Γ is known and $r_t^m$is specified, $G_t$is fully observable and a completely tractable model of mean-reverting relative pricing for two stocks A and B exists, which can be used for pairs trading strategies (Fiz, 2014).

Unlike pairs trading strategies which are predicted on mispricing at the price level, the strategy proposed by Do *et al.* (2006) is based on mispricing at return levels. The practical implication is that the proposed strategy opens positions when the accumulated residual spread in the returns is sufficiently large, and unwind when the accumulated spread is equal to the long run level of spread, in contrast to the other models in which positions open when the prices drift sufficiently apart and unwind when they converge (Do *et al.*, 2006).

1.5. Stochastic Control Approach

Jurek and Yang (2007) are the most cited authors in this domain. Their model assumes that the size of the positions that arbitrageurs are willing to take is affected by two kinds of risk: the horizon risk and the divergence risk. The former represents the uncertainty about the timing in which the mispricing will be eliminated, the other one represents the uncertainty about a possible deterioration in the mispricing prior to its elimination.

The arbitrage opportunity, which is interpretable as a long/short relative value trade whose magnitude is measured by the price differential (spread), is described by a mean-reverting process (Ornstein-Uhlenbeck process) in order to capture these two forms of risk.

Under this assumption, the arbitrageur faces uncertainty about the magnitude of mispricing at all future date. In Jurek and Yang's model (2007) arbitrageurs can invest in a riskless bond (B_t) and in the mean-reverting spread (S_t) whose dynamics are described by:

$$dB_t = rB_t dt$$

$$dS_t = \kappa(\bar{S} - S_t)dt + \sigma dZ_t$$

If $S_t > \bar{S}$, meaning that at time t the spread is higher than the long-run mean spread ($\bar{S}$), the arbitrageurs shorts the spread (buying the undervalued and selling the overvalued) and invests the revenues in the risk-free assets. If the reverse is true, the agent goes long on the spread and invests the proceeds in the riskless asset.

Jurek and Yang (2007) derive the arbitrageur's optimal dynamic portfolio policy for two different non-myopic preference specifications:

- Constant relative risk aversion (CRRA) defined over final wealth at a finite horizon
- Epstein-Zin recursive utility defined over intermediate consumption

Denoting with N_t and M_t respectively the number of units of the spread and of the bond held by the arbitrageurs it is possible to compute the budget constraints for both specifications:

$$\begin{cases} dW_t = N_t dS_t + M_t dB_t & \text{for CRRA specification} \\ dW_t = N_t dS_t + M_t dB_t - C_t dt & \text{for Epstein} - \text{Zin specification} \end{cases}$$

where W_t represents the wealth of the investor at time t and C_t represents the consumption rate, which affects the evolution of wealth only when the Epstein-Zin form is considered. Substituting the assets' prices dynamics into the budget constraints one obtains:

$$\begin{cases} dW_t = (r(W_t - N_t S_t) + \kappa(\bar{S} - S_t)N_t)dt + \sigma N_t dZ_t & \text{for CRRA specification} \\ dW_t = (r(W_t - N_t S_t) + \kappa(\bar{S} - S_t)N_t - C_t)dt + \sigma N_t dZ_t & \text{for Epstein} - \text{Zin specification} \end{cases}$$

Given these budget constraints, Jurek and Yang (2007) applying the standard stochastic control theory derive the Hamilton-Jacobi-Bellmann (HJB) equation for each stochastic dynamic programming problem and find closed-form solutions for the policy and value functions, which are the optimal strategy as a function of the state variables of the model, and the best possible value of the objective function expressed as a function of the state variables, respectively.

Another important finding of the study by Jurek and Yang (2007) is that arbitrageurs do not always perform arbitrage. The authors analytically demonstrate that there is a crucial level of mispricing beyond which further divergence can result in a decline of the portion of wealth allocated to the spread portfolio.

According to classic economic theory as the spread widens the arbitrageur increases the proportion of wealth invested in the spread portfolio, thus producing a stabilizing effect on the mispricing and contributing to its elimination. In practice, it is also possible that the arbitrageur reduces his position in response to an adverse shock, meant as a change in the market conditions which has the effect of distancing the spread from its long-run equilibrium value, thus producing a destabilizing effect and exacerbating the mispricing. The direction in which the agent trades in response to adverse shocks to the value of the spread depends on the combination of two different effects: the wealth effect and the 'investment opportunity' effect.

An adverse shock, since it increases the magnitude of the deviation of the spread from its long-term mean, makes the investment opportunity more attractive, inducing the investors to take larger positions in the spread portfolio relative to their wealth, but at the same time it has a negative effect on the arbitrageur's wealth since it causes traders to lose money on their current positions, leading them to reduce their positions.

As long as the improvement in the investment opportunity dominates the wealth effect, the arbitrageur will increase his position in the mean-reverting portfolio, while if the wealth effect outweighs the 'investment opportunity' effect the agent will reduce his position in the spread portfolio.

Analytically it is possible to identify a stabilization region: if the spread is within this stabilization region a rational agent will increase his position in the spread asset in response to increasing mispricing, if not he will reduce his position thus exacerbating the mispricing (Jurek and Yang, 2007).

Based on the work by Jurek and Yang (2007) Liu and Timmermann (2013), using a cointegration framework for the asset price dynamics, show that the optimal convergence trading strategy that maximizes expected utility generally does not involve holding a delta neutral long-short position. Delta neutral trades aimed at exploiting temporary mispricing between assets by taking long-short positions in such a way that the market exposure gets eliminated. According to Liu and Timmerman (2013), the main limitation of the delta neutral approach, that is also the reason why this approach cannot be the most efficient way to exploit a temporary mispricing, is that it does not consider the trade-off between diversification and arbitrage. By focusing only on long-term arbitrage, delta neutral strategies do not completely exploit the short-term risk-return trade-off and diversification benefits. By examining the arbitrage opportunity in the context of a portfolio maximization problem, the strategy developed by Liu and Timmerman (2013), accounts for both arbitrage opportunities and diversification benefits.

Following Jurek and Yang (2007), the authors derive the HJB equation for an investor maximizing the expected value of a power utility function defined over terminal wealth. They find the value and policy functions from which the optimal portfolio weights can be derived. Liu and Timmermann (2013) obtain two surprising results: first, it can be optimal to hold both risky assets long (or short) at the same time, even if prices eventually converge (only in multiperiod model). Second, it can be optimal to hold just one asset disregarding the second. This optimal investment policy is in stark contrast to standard delta neutral long-short strategy.

1.6. Copula Approach

The copula approach has been conceived in order to overcome one of the main issues related to the most commonly used pairs trading techniques (distance approach and cointegration approach) that is the assumption of linear dependence and the relative use of correlation coefficient or cointegration as measures of dependency. This assumption is too simplistic, real financial data are very rarely normally distributed, therefore correlation and cointegration cannot completely describe the dependency and predict the future movements (Liew and Wu, 2013).

A copula establishes a functional relationship between a multivariate distribution function and its marginals; it captures the dependence structure between the marginal distributions (Rad *et al.*, 2016). The copula approach can solve the problems mentioned earlier as it uses a two-step methodology that separates the choice of the best-fitting marginal distribution describing the variables from the application of a suitable copula to establish the dependence.

The use of copulas produces greater flexibility in the framework when joint distributions are specified, while providing richer information regarding the dependency between stocks (Liew and Wu, 2013).

From an analytical point of view, following Krauss and Stübinger (2015), any function *C:* $[0,1]^n \rightarrow [0,1]$ is an n-dimensional copula if three conditions are satisfied;

- $\forall u = (u_1, u_2, \ldots, u_n) \in [0,1]^n: min\{u_1, u_2, \ldots, u_n\} = 0 \rightarrow C(u) = 0,$
- $C(1, \ldots, 1, u_i, 1, \ldots, 1) = u_i \; \forall \; u_i \in [0,1] \; (i \in 1, \ldots, n),$
- $V_C([a, b]) \geq 0$, where $V_C([a, b])$ denotes the C-volume of the hyperrectangle

  $$[a, b] = \prod_{i=1}^{n} [a_i, b_i], a_i \leq b_i \quad \forall i \in 1, \ldots, n$$

As expressed in Sklar's theorem: let $F_{X_1,\ldots,X_n}$ be an n-dimensional distribution function with marginal distributions F_{X_i}(i=1,…,n). Then, there exists an n-copula *C* which satisfies the following equation for all $(x_1, \ldots, x_n) \in R^n$:

$$F_{X_1,\ldots,X_n}(x_1, \ldots, x_n) = C\left(F_{X_1}(x_1), \ldots, F_{X_n}(x_n)\right)$$

If the margins are continuous, then *C* is unique (Sklar, 1959).

According to Krauss (2017), two sub-streams can be found in the literature: the return-based copula approach and the level-based copula approach.

Return-based copula approach: during the formation period, pairs are selected applying either correlation criteria or cointegration criteria. Then, the log returns $R_1 = (R_{1,t})_{t \in T}$ and $R_2 = (R_{2,t})_{t \in T}$ for the two stocks of each pairs are considered, as well as their marginal distributions F_{R_1} and F_{R_2}, which can be estimated either using fitting parametric distribution functions (Liew and Wu, 2013) or parametric and non-parametric approaches (Stander *et al.*, 2013). At this point, applying probability integral transform by plugging the returns into their distribution functions creates two uniform variables, $U_1 = F_{R_1}(R_1)$ and $U_2 = F_{R_2}(R_2)$, which allow the identification of an appropriate copula function (Krauss, 2017). Several methodologies can be used for the identification of the copula function: Stander *et al.* (2013) rely on a set of 22 Archimedean copulas, listed by Nelsen (2006), and determine the best-fit applying the Kolmogorov-Smirnov goodness-of-fit test, while Liew and Wu (2013) use five copula largely applied in financial applications (Gumbel, Student-*t*, Normal, Frank and Clayton) and select the best fitting one using three different information criteria (Schwarz Information Criteria, Akaike Information Criterion and Hannan-Quinn Information Criterion). At this point, the best fitting copula is used in order to calculate the conditional marginal distribution functions as first partial derivatives of the copula function *C* (u_1, u_2):

$$h_1(u_1|u_2) = P(U_1 \leq u_1|U_2 = u_2) = \frac{\partial C(u_1, u_2)}{\partial u_2}$$

$$h_2(u_2|u_1) = P(U_2 \leq u_2|U_1 = u_1) = \frac{\partial C(u_1, u_2)}{\partial u_1}$$

Stocks are identified as relatively undervalued if the conditional probability is less than 0.5 and relatively overvalued if the conditional probability is greater than 0.5.

Stander *et al.* (2013) and Liew and Wu (2013) suggest trading when the conditional probabilities are in the tail regions of the conditional distribution functions.

Placing the upper bound at 0.95 and the lower bound at 0.05 for the threshold of conditional probabilities, trading occurs if the pair of transformed returns falls outside both confidence band derived by $P(U_2 \leq u_2|U_1 = u_1)$= 0.05 and $P(U_1 \leq u_1|U_2 = u_2)$=0.95; in this particular case stock 1 is sold and stock 2 is bought. The opposite transactions will occur if inverse conditions apply.

For what concerns the exit strategies: Liew and Wu (2013) suggest exiting a trade once the positions revert (i.e. when the conditional probabilities cross the boundary of 0.5); Stander *et al.* reverse their positions as soon as it is profitable or alternatively after one trading week, since pairs trading is a short-term strategy.

The main issues with this approach are: first, pairs selection is not copula-based, and this introduces a selection bias. Second, the time structure of data is completely lost, meaning that copula-based entry and exit signals are anchored on the last return without assessing how each pair trades after the entry signal (Krauss and Stübinger, 2015).

Level-based copula method: this method is introduced by Rad *et al.* (2016) and Xie *et al.* (2014). According to Rad *et al*. (2016), the trading strategy is the following: first, during a formation period 20 pairs with the least SSD are nominated for trading during the trading period (like Gatev *et al.*, 2006). For each pair, the daily returns of the formation period are fitted to the best-fitting marginal distributions and by maximizing the log likelihood of each copula density function the best fitting copula is nominated.

At this point, using the daily realization of random variables U_1 and U_2 it is possible to calculate the conditional probabilities h_1 and h_2 for each pair used, similarly to Xie *et al.* (2014), to define two mispriced indices:

$$m_{1,t} = h_{1,t}(u_{1,t}|u_{2,t}) - 0.5 = P(U_{1,t} \leq u_{1,t}|U_{2,t} = u_{2,t}) - 0.5$$

$$m_{2,t} = h_{2,t}(u_{2,t}|u_{1,t}) - 0.5 = P(U_{2,t} \leq u_{2,t}|U_{1,t} = u_{1,t}) - 0.5$$

Positive value of $m_{1,t}$ and negative value of $m_{2,t}$ can be interpreted as stock 1 being overvalued with respect to stock 2 at time *t*, and vice versa.

Consider the cumulative mispricing indices $M_{1,t}$ and $M_{2,t}$, set to zero at the beginning of the trading period and calculated each day as:

$$M_{1,t} = M_{1,t-1} + m_{1,t}$$

$$M_{2,t} = M_{2,t-1} + m_{2,t}$$

Positive (negative) M_1 and negative (positive) M_2 are interpreted as stock 1 (stock 2) being overvalued with respect to stock 2 (stock 1).

The main downsides of the level-based copula approach are: first, pairs selection is not copula-based, the 20 pairs with minimum SSD are selected introducing a severe selection bias. Second, there is no differentiation between pairs reaching critical levels of mispricing indices over time by aggregating many small mispricings versus pairs reaching the critical levels in a few large steps (Krauss and Stübinger, 2015).

Chapter 2

Cointegration Approach: Statistical Tools

This chapter is dedicated to the introduction of the elements necessary for understanding the concept of cointegration, as exposed by Engle and Granger in 1987. Section 2.1 briefly introduces a very useful operator in time series analysis, that is the *lag operator*. Section 2.2 presents the definition of two crucial concepts used in time series analysis, which are *stationarity* and *ergodicity*. Section 2.3 and 2.4 investigate the concept and the most relevant properties of univariate *ARMA* processes. Section 2.5 and 2.6 analyse different type of nonstationary time series and discuss the different techniques which are used to produce stationary time series. Section 2.7 and 2.8 present two different types of test used to determine if a time series variable is nonstationary and possesses a unit root. Section 2.9 explores an important problem, known as spurious regression, that can arise if the error terms in a regression are integrated of order one. Section 2.10 introduces the concept of cointegration and develops different representations of a cointegrating system. Section 2.11 examines two different tests used to detect cointegration among the elements of one or more time series: the first is a simple residual-based testing method, while the second is based on full-information maximum likelihood estimation.

2.1 Lag Operator

A time series is a collection of observations indexed by the date of each observations. The collected data begin at some particular date (*t=1*) and end at another date (*t=T*):

$$(y_1, y_2, y_3, \ldots, y_T)$$

Statisticians often imagine that they could have obtained earlier observations $(y_0, y_{-1}, y_{-2}, \ldots)$ or later observations $(y_{T+1}, y_{T+2}, y_{T+3}, \ldots)$ had the process been observed for more time. The observed sample $(y_1, y_2, y_3, \ldots, y_T)$ could be view as a finite segment of a doubly infinite sequence, denoted $\{y_t\}_{t=-\infty}^{\infty}$:

$$y_t = \{\ldots, y_{-1}, y_0, \underbrace{y_1, y_2, \ldots, y_T}_{\text{observed sample}}, y_{T+1}, y_{T+2}, \ldots\}$$

A time series operator transforms one time series or group of time series into a new time series. It takes as input a sequence such as $\{k_t\}_{t=-\infty}^{\infty}$ or a group of sequences such as $(\{z_t\}_{t=-\infty}^{\infty}, \{w_t\}_{t=-\infty}^{\infty})$ and it gives as output a new sequence $\{y_t\}_{t=-\infty}^{\infty}$.

A useful operator, used in time series analysis, is the lag operator (L) which operates on an element of a time series to produce the previous element. Consider a sequence $\{k_t\}_{t=-\infty}^{\infty}$ which generates a new sequence $\{y_t\}_{t=-\infty}^{\infty}$, where the value of *y* for date *t* is equal to the value of *k* at date *t-1*:

$$y_t = k_{t-1}$$

This can be described applying the lag operator to the sequence $\{k_t\}_{t=-\infty}^{\infty}$:

$$Lk_t \equiv k_{t-1} = y_t$$

From the previous equation, it follows that:

$$L^2 k_t = L(Lk_t) = Lk_{t-1} = k_{t-2}$$

In general, for any integer number *p* and *q*:

$$L^p k_t = k_{t-p}$$

$$(L^p)^q k_t = L^{pq} k_t = k_{t-pq}$$

$$(L^p)(L^q)k_t = L^p k_{t-q} = L^{p+q} k_t = k_{t-p-q}$$

2.2 Stationarity and Ergodicity

A time series $\{y_t\}_{t=-\infty}^{\infty}$ is said to be *weakly stationary* (or *covariance-stationary*) if neither its mean μ_t nor its autocovariances γ_{jt} depend on the date *t*, that is:

$$E(y_t) = \mu \qquad \text{for all } t$$

$$Cov(y_t, y_{t-j}) = E[(y_t - \mu)(y_{t-j} - \mu)] = \gamma_j \qquad \text{for all } t \text{ and any } j$$

If a process is weakly stationary its mean and variance (γ_0) are constant and finite over time, and the covariances between any two observations of the process (e.g. y_t and y_{t-j}) depends only on the length of time separating the two observations, and not on the date of the observation (*t*). It follows that for a covariance-stationary process, γ_j and γ_{-j} would represent the same quantity (Hamilton, 1994, pp.45-46).

A time series $\{z_t\}_{t=-\infty}^{\infty}$ is *strictly stationary* if the joint probability distribution of any set of *m* consecutive observations $[z_t, z_{t+1}, \ldots, z_{t+m-1}]$ is the same regardless of the time instant *t* (Greene, 2018, p. 992). In other words, a process is strictly stationary if the distribution of its value remains the same as time progresses, implying that the probability that z_t increases or falls within a particular interval is the same at any time in the past or in the future (Brooks, 2008, p. 208).

Notice that if a process is strictly stationary with finite second moments, then it must be weakly stationary. However, it is possible that a weakly stationary process is not strictly stationary, because the mean and the autocovariances are time-invariant but higher moments such as $E(y_t^3)$ or $E(y_t^4)$ are not (Hamilton, 2004, p. 46).

For the purpose of this thesis only weak stationarity is required so from now on, the term *stationarity* (or *stationary*) by itself is taken to mean covariance stationarity.

Another crucial property of time series that needs to be examined is *ergodicity*. Consider the following sample of T observations from the process $\{y_t\}_{t=-\infty}^{\infty}$,denoted $\{y_1, y_2, .., y_T\}$.From these observations is possible to calculate the sample mean $\bar{y}$, that is:

$$\bar{y} = \frac{1}{T}\sum_{t=1}^{T} y_t$$

A weakly stationary process is ergodic for the mean if the sample mean converges in probability to the unconditional mean of the process $E(y_t) = \mu$, as $T \to \infty$. In other words, the sample mean provides an unbiased estimate of the population mean.

Applying the law of large numbers to a covariance stationary process it is possible to demonstrate that if the autocovariances satisfies the following condition:

$$\sum_{j=0}^{\infty} |\gamma_j| < \infty$$

then the process $\{y_t\}_{t=-\infty}^{\infty}$ is ergodic for the mean (Hamilton, 1994, pp. 46-47).

2.3 White Noise Processes

A process $\{\varepsilon_t\}_{t=-\infty}^{\infty}$ is described as a *white noise process* if it satisfies the following condition:

$$E(\varepsilon_t) = \mu \tag{2.1}$$

$$Var\,(\varepsilon_t) = \sigma^2 \tag{2.2}$$

$$E(\varepsilon_t, \varepsilon_s) = 0 \quad \text{for } t \neq s \tag{2.3}$$

Thus, a white noise process is a process with mean and variance that are constant and finite over time and for which the random variables ε are uncorrelated across time. In some circumstances it could be useful to replace the condition regarding the autocorrelation with a stronger one, that is:

$$\varepsilon_t, \varepsilon_s \ are\ indendent\ for\ t \neq s$$

A process with mean and variance that are constant over time and for which the random variables ε are independent is called an *independent white noise*. The difference between the white noise process and the independent white noise process is that the first has uncorrelated increments, while the latter has independent increments. Finally, an independent white noise process for which

$$\varepsilon_t \sim N(0, \sigma^2)$$

is defined as a *Gaussian white noise process* (Hamilton, 1994, pp. 47-48).

2.4 Stationary *ARMA* Processes

This section presents the definition and the notation of univariate Autoregressive Moving Average (*ARMA*) processes, which provide a valuable class of models for time series analysis. In order to understand the concept of Autoregressive Moving Average process, it is essential to introduce the definition and notation of two key time series models which are the Moving Average (*MA*) process and the Autoregressive (*AR*) process.

The qth-Order and Infinite Order Moving Average Process

A *qth-order Moving Average process*, denoted *MA*(q), is characterized by (Hamilton, 1994, p.50):

$$y_t = \mu + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \ldots + \theta_q \varepsilon_{t-q} \qquad (2.4)$$

where $\{\varepsilon_t\}_{t=-\infty}^{\infty}$ satisfies the conditions (2.1), (2.2) and (2.3), $(\theta_1, \theta_2, \ldots, \theta_q)$ could be any real numbers and μ is the mean of the process and could be any real number. The term "moving average" comes from the fact that the process is constructed from a weighted sum of the q most recent values of ε. The mean, variance and autocovariances of the process can be computed as follow:

$$E(y_t) = \mu + E(\varepsilon_t) + \theta_1 E(\varepsilon_{t-1}) + \theta_2 E(\varepsilon_{t-2}) + \ldots + \theta_q E(\varepsilon_{t-q}) = \mu$$

$$Var(y_t) = E(y_t - \mu)^2 = E(\varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \ldots + \theta_q \varepsilon_{t-q})^2 = (1 + \theta_1^2 + \theta_2^2, \ldots + \theta_q^2)\sigma^2$$

For $j = 1, 2, \ldots, q$,

$$\gamma_j = E[(\varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \ldots + \theta_q \varepsilon_{t-q})(\varepsilon_{t-j} + \theta_1 \varepsilon_{t-j-1} + \theta_2 \varepsilon_{t-j-2} + \ldots + \theta_q \varepsilon_{t-j-q})]$$

$$= E[\theta_j \varepsilon_{t-j}^2 + \theta_{j+1}\theta_1 \varepsilon_{t-j-1}^2 + \theta_{j+2}\theta_2 \varepsilon_{t-j-2}^2 + \ldots + \theta_q \theta_{q-j} \varepsilon_{t-q}^2]$$

$$\gamma_j = \begin{cases} (\theta_j + \theta_{j+1}\theta_1 + \theta_{j+2}\theta_2 + \ldots + \theta_q \theta_{q-j})\sigma^2 & \text{for } j = 1, 2, \ldots, q \\ 0 & \text{for } j > q \end{cases}$$

Notice that the mean and the variance of the *MA*(q) process are finite and constant over time, and the covariances between any two observations of the process depends only on the length of time separating the two observations; thus, the *MA*(q) process is covariance stationary for any value of $(\theta_1, \theta_2, \ldots, \theta_q)$.

The *infinite-order Moving Average process*, denoted *MA* (∞), is the process resulting from a *MA*(q) process when $q \to \infty$ (Hamilton, 1994, p. 52):

$$y_t = \mu + \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j} = \mu + \psi_0 \varepsilon_t + \psi_1 \varepsilon_{t-1} + \psi_2 \varepsilon_{t-2} + \ldots.$$

It is possible to demonstrate that this infinite sequence generates a well-defined covariance-stationary process provided that the sequence $\{\psi_j\}_{j=0}^{\infty}$ is absolutely summable, that is (see Hamilton, 1994, pp. 69-70):

$$\sum_{j=0}^{\infty} |\psi_j| < \infty$$

where ψ_j represents the coefficients of an infinite order moving average process.

The pth-Order Autoregressive Process

A *pth-order Autoregressive process*, denoted *AR*(p), is characterized by (Hamilton, 1994, p. 58):

$$y_t = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \ldots + \phi_p y_{t-p} + \varepsilon_t \qquad (2.5)$$

where $\{\varepsilon_t\}_{t=-\infty}^{\infty}$ satisfies the conditions (2.1), (2.2) and (2.3), c is a constant which could be any real number. This difference equation is stable, in the sense that the consequences of any shock gradually die out, provided that all the roots of the following characteristic equation:

$$1 - \phi_1 z - \phi_2 z^2 - \ldots - \phi_p z^p = 0$$

lie outside the unit circle. When this condition is satisfied, the *AR*(p) process turns out to be covariance-stationary, and the inverse of the autoregressive operator is:

$$\psi(L) = (1 - \phi_1 L - \phi_2 L^2 - \ldots - \phi_p L^p)^{-1} = \psi_0 + \psi_1 L + \psi_2 L^2 + \ldots.$$

with $\sum_{j=0}^{\infty} |\psi_j| < \infty$. Multiplying both sides of equation (2.5) by $\psi(L)$ gives:

$$y_t = \mu + \psi(L)\varepsilon_t$$

which can be viewed as a *MA* (∞) process with mean $\mu = c/(1 - \phi_1 - \ldots - \phi_p)$.

Equation (2.5) can be rewritten as:

$$y_t - \mu = \phi_1(y_{t-1} - \mu) + \phi_2(y_{t-2} - \mu) + \ldots + \phi_p(y_{t-p} - \mu) + \varepsilon_t$$

from which is possible to find the autocovariances and the variance (γ_0) of the process simply by multiplying both sides of the equation by $(y_{t-j} - \mu)$ and taking the expectation, such that:

$$\gamma_j = \begin{cases} \phi_1 \gamma_{j-1} + \phi_2 \gamma_{j-2} + \ldots + \phi_p \gamma_{j-p} & \text{for } j = 1, 2, \ldots \\ \phi_1 \gamma_1 + \phi_2 \gamma_2 + \ldots + \phi_p \gamma_p + \sigma^2 & for\ j = 0 \end{cases}$$

The Autoregressive Moving Average Process

An *Autoregressive Moving Average process*, denoted *ARMA*(*p*, *q*), includes both autoregressive and moving average terms, and it is characterized by (Hamilton, 1994, p. 59):

$$y_t = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \ldots + \phi_p y_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \ldots + \theta_q \varepsilon_{t-q} \quad (2.6)$$

which can be rewritten in lag form as:

$$y_t(1 - \phi_1 L - \phi_2 L^2 - \ldots - \phi_p L^p) = c + (1 + \theta_1 L + \theta_2 L^2 + \ldots + \theta_q L^q)\varepsilon_t \quad (2.7)$$

If all the roots of the characteristic equation:

$$1 - \phi_1 z - \phi_2 z^2 - \ldots - \phi_p z^p = 0$$

lie outside the unit circle, both sides of the equation (2.7) can be divided by $(1 - \phi_1 L - \phi_2 L^2 - \ldots - \phi_p L^p)^{-1}$ to get:

$$y_t = \mu + \psi(L)\varepsilon_t$$

with $\sum_{j=0}^{\infty} |\psi_j| < \infty$ and where $\psi(L)$ and μ are respectively:

$$\psi(L) = \frac{(1 + \theta_1 L + \theta_2 L^2 + \ldots + \theta_q L^q)}{(1 - \phi_1 L - \phi_2 L^2 - \ldots - \phi_p L^p)} \qquad \mu = \frac{c}{(1 - \phi_1 - \phi_2 - \ldots - \phi_p)}$$

Thus, from the equation describing μ it is easy to see that the stationarity of an *ARMA* process depends entirely on the autoregressive parameters $(\phi_1, \phi_2, \ldots, \phi_p)$ and not on the moving average parameters $(\theta_1, \theta_2, \ldots, \theta_q)$.

Invertibility for the qth-Order Moving Average Process

Consider the following *MA*(*q*) process:

$$(y_t - \mu) = (1 + \theta_1 L + \theta_2 L^2 + \ldots + \theta_q L^q)\varepsilon_t \quad (2.8)$$

where $\{\varepsilon_t\}_{t=-\infty}^{\infty}$ satisfies the condition (2.1), (2.2) and (2.3). Provided that all the roots of the characteristic equation:

$$1 - \theta_1 z - \theta_2 z^2 - \ldots - \theta_q z^q = 0$$

lie outside the unit circle, then equation (2.8) can be written as an *AR*(∞) by inverting the *MA* operator (Hamilton, 1994, p. 67):

$$(1 + \theta_1 L + \theta_2 L^2 + \ldots + \theta_q L^q)^{-1}(y_t - \mu) = \varepsilon_t$$

with:

$$(1 + \theta_1 L + \theta_2 L^2 + \ldots + \theta_q L^q)^{-1} = (1 + \eta_1 L + \eta_2 L^2 + \ldots.)$$

If these conditions are satisfied, the *MA*(*q*) process, described in equation (2.8), is invertible.

2.5 Integrated Processes

Consider the following process:

$$y_t = \alpha + \delta t + u_t \qquad \text{for all } t \tag{2.9}$$

where u_t follows a zero-mean *ARMA* (p, q) process (see Section 2.4) described by the following equation:

$$u_t = \phi_1 u_{t-1} + \phi_2 u_{t-2} + \ldots + \phi_p u_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \ldots + \theta_q \varepsilon_{t-q} \tag{2.10}$$

where $\{\varepsilon_t\}_{t=-\infty}^{\infty}$ is a white noise sequence with mean zero and variance σ^2 and $(\theta_1, \theta_2, \ldots, \theta_q)$ could be any real number. Equation (2.10) can be rewritten in lag form as:

$$(1 - \phi_1 L - \phi_2 L^2 - \ldots - \phi_p L^p) u_t = (1 + \theta_1 L + \theta_2 L^2 + \ldots + \theta_q L^q)\, \varepsilon_t \tag{2.11}$$

where the moving average operator $(1 + \theta_1 L + \theta_2 L^2 + \ldots + \theta_q L^q)$ is invertible (see Section 2.4). Consider the following factorization of the autoregressive operator (see Hamilton, 1994, p. 33):

$$(1 - \phi_1 L - \phi_2 L^2 - \ldots - \phi_p L^p) = (1 - \lambda_1 L)(1 - \lambda_2 L)\ldots(1 - \lambda_p L)$$

In order to solve this equation, it is necessary to substitute the lag operator L with a scalar z because L denotes a particular operator, not a number, and find a result for L would not be a sensible statement:

$$(1 - \phi_1 z - \phi_2 z^2 - \ldots - \phi_p z^p) = (1 - \lambda_1 z)(1 - \lambda_2 z)\ldots(1 - \lambda_p z) \tag{2.12}$$

The goal is to find the values of $(\lambda_1, \lambda_2, \ldots, \lambda_p)$ so that the two sides of the equation (2.12) represent the identical polynomial in z. It is easy to see that the left hand side is equal to zero if $z \equiv \lambda^{-1}$, so if one finds a value of z that sets the right hand side to zero, that value of z must set the left hand side to zero as well.

If all the values $(\lambda_1, \lambda_2, \ldots, \lambda_p)$ are inside the unit circle, that is $|\lambda_i| < 1$, then equation (2.11) represents a stationary process and can be expressed as:

$$u_t = \frac{(1 + \theta_1 L + \theta_2 L^2 + \ldots + \theta_q L^q)}{(1 - \lambda_1 L)(1 - \lambda_2 L)\ldots(1 - \lambda_p L)} \varepsilon_t = \psi(L)\varepsilon_t$$

with absolute summability of the moving average coefficients (see Section 2.4), that is $\sum_{j=0}^{\infty} |\psi_j| < \infty$, and the roots of the characteristic equation $\psi(z) = (1 - \phi_1 z - \phi_2 z^2 - \ldots - \phi_p z^p) = 0$ outside the unit circle.

Consider the case in which $\lambda_1 = 1$ and $|\lambda_i| < 1$ for i =2, 3,…, p. From (2.11) one obtains:

$$(1 - L)(1 - \lambda_2 L)\ldots(1 - \lambda_p L) u_t = (1 + \theta_1 L + \theta_2 L^2 + \ldots + \theta_q L^q)\, \varepsilon_t$$

this implies that:

$$(1-L)u_t = \frac{(1+\theta_1 L + \theta_2 L^2 + \ldots + \theta_q L^q)}{(1-\lambda_2 L)(1-\lambda_3 L)\ldots(1-\lambda_p L)}\ \varepsilon_t$$

with $\sum_{j=0}^{\infty}|\psi_j| < \infty$ and the roots of $\psi(z) = (1-\phi_1 z - \phi_2 z^2 - \ldots - \phi_p z^p) = 0$ outside the unit circle. At this point, if (2.9) is first-differenced, the result is:

$$(1-L)y_t = (1-L)\alpha + [\delta t - \delta(t-1)] + (1-L)u_t$$

$$= 0 + \delta + \frac{(1+\theta_1 L + \theta_2 L^2 + \ldots + \theta_q L^q)}{(1-\lambda_2 L)(1-\lambda_3 L)\ldots(1-\lambda_p L)}\ \varepsilon_t$$

which is a *unit root process*. If a process written in the form (2.9) and (2.11) has one eigenvalue λ_1 equals to one and all the others inside the unit circle, then the first difference produces a stationary process. This process is said to be integrated of order one, denoted *I*(1) (Hamilton, 1994, p. 437). In general, a nonstationary time series is integrated of order *d*, denoted *I*(*d*), if it becomes stationary, denoted *I*(0), after being differenced *d* times, which is equivalent to say that the process has *d* eigenvalues equal to one and all the other inside the unit circle.

A process written in the form of (2.9) and (2.11) is a generalization of an *ARMA* model (see Section 2.4), which is called *autoregressive integrated moving average process*, denoted *ARIMA* (*p*, *d*, *q*), where the first parameter (*p*) refers to the number of time lags of the autoregressive model (not including the unit roots), the second parameter (*d*) refers to the order of integration and the third parameter (*q*) refers to the order of the moving average model. Taking *d* differences of an *ARIMA* (*p, d, q*) process produces a stationary *ARMA* (*p, q*) process.

2.6 Non-Stationarity: Random Walk with and without drift and Trend-Stationary Processes

There are two important models that have been used in order to represent the non stationarity:

- The random walk with drift:

$$y_t = \mu + y_{t-1} + \varepsilon_t \text{ with } \varepsilon_t \sim i.i.d.(0, \sigma^2) \tag{2.13}$$

  where μ is a constant representing the drift in the process.

- The trend-stationary process:

$$y_t = \alpha + \beta t + \varepsilon_t \text{ with } \varepsilon_t \sim i.i.d.(0, \sigma^2) \tag{2.14}$$

The random walk with drift model can be generalized as follow:

$$y_t = \mu + \phi y_{t-1} + \varepsilon_t$$

where ϕ is the coefficient of the term y_{t-1} and could be any real number.

This generalization is useful to understand how a shock at time (*t-T*) will affect the system in the future (*t*). For this purpose, it is sufficient to consider an *AR* (1) model with no drift:

$$y_t = \phi y_{t-1} + \varepsilon_t \tag{2.15}$$

Lagging the equation (2.15), one obtains:

$$y_{t-1} = \phi y_{t-2} + \varepsilon_{t-1} \tag{2.16}$$
$$y_{t-2} = \phi y_{t-3} + \varepsilon_{t-2} \tag{2.17}$$

Substituting (2.16) and (2.17) into (2.15) the result is:

$$y_t = \phi^3 y_{t-3} + \phi^2 \varepsilon_{t-2} + \phi \varepsilon_{t-1} + \varepsilon_t$$

Repeating this substitution T times leads to the following result:

$$y_t = \phi^{T+1} y_{t-T-1} + \phi \varepsilon_{t-1} + \phi^2 \varepsilon_{t-2} + \ldots + \phi^T \varepsilon_{t-T} + \varepsilon_t$$

At this point, three different scenarios are possible (Brooks, 2008, pp. 321-322):

- If $\phi < 1 \; then \; \phi^T \to 0 \; as \; T \to \infty$: in this case a shock to the system gradually disappears (*stationary case*).
- If $\phi = 1 \; then \; \phi^T = 1 \; \forall T$: in this case a shock persists in the system (*unit root case*).
- If $\phi > 1$: in this case a shock to the system becomes more influential as time goes on (*explosive case*).

The random walk with drift described in (2.13) can be written also as:

$$y_t - y_{t-1} = (1 - L) y_t = \mu + \varepsilon_t$$

The characteristic equation associated to the previous equation is (see Section 2.5):

$$1 - z = 0$$

And the root of this characteristic equation is one, meaning that the process is integrated of order one, thus nonstationary. Indeed, it is easy to see that the first difference of $\{y_t\}_{t=-\infty}^{\infty}$:

$$w_t = \Delta y_t = y_t - y_{t-1} = \mu + \varepsilon_t \tag{2.18}$$

is simply the sum of the innovation (ε_t) and the mean of the process $\{w_t\}_{t=-\infty}^{\infty}$, and so it is a stationary process. Thus, the random walk with drift is an integrated process of order one, denoted $y_t \sim I(1)$.

The trend-stationarity process is composed by a trend process ($\alpha + \beta t$) plus a stationary process represented by the innovation, $\varepsilon_t \sim I(0)$.

Both processes will produce strongly trended nonstationary series, the difference is that the random walk has a *stochastic trend*, implying that the series will increase in each period by a stochastic amount given by the realization of the disturbance, while the trend-stationary process will increase in each period by a deterministic amount, due to the *deterministic trend*, plus the concrete realization of the random disturbance. The two processes are very different from each

other and when analysing a phenomenon is crucial to understand which process is generating such phenomenon because the transform required to make the process stationary is not the same for the two models (Alexander, 2008, pp. 214-215).
To transform a random walk, or any integrated process of order one, into a stationary process it is sufficient to take the *first difference*, as in equation (2.18). In order to transform a trend-stationarity process into a stationary process *detrending* is required; in other words, a regression in the form given in (2.14) would be run, and any subsequent estimation would be done on the residuals from (2.14), which would have had the linear trend removed (Brooks, 2008, p. 322).

2.7 Unit Root Tests

Statistical tests of the null hypothesis that a time series is non-stationary versus the alternative that it is stationary are called *unit root tests* from the fact that an autoregressive series is stationary, or integrated of order zero, if and only if the roots of its characteristic equation lie inside the unit circle. The hypotheses for a unit root test are the following:

$$H_0: y_t \sim I(1) \quad vs \quad H_1: y_t \sim I(0)$$

Many economic and financial time series exhibit trending behaviour (i.e. nonstationary in the mean), and as discussed in Section 2.6 a crucial econometric task is determining the most appropriate form of the trend that is present in the data, because there exists different trend removal procedures (first differencing and detrending). The role of unit root tests is crucial in determining whether the data are integrated processes of order one which need to be first differenced or trend-stationary processes which need to be detrended (Zivot and Wang, 2006, p. 107).
The major shortcoming of the unit root tests is their low statistical power against $I(0)$ alternatives that are close to be $I(1)$. In other terms, the test is unable to distinguish between unit root and near-unit root time series. With near unit root time series, the risk is to be unable to reject the null hypothesis of unit root. This means that the test has a high probability of not rejecting a false null hypothesis (Kočenda and Černý, 2015, p. 72).

2.7.1 Dickey-Fuller test

The early and pioneering work on testing for unit root in time series was done by Dickey and Fuller (1979). The Dickey-Fuller (DF) test allows to evaluate the presence of a trend or of unit roots in the series analysed.
Consider the following autoregressive model:

$$y_t = \phi y_{t-1} + u_t \tag{2.19}$$

Where $y_0 = 0$, ϕ is the coefficient of the term y_{t-1} and could be any real number, $u_t \sim N(0, \sigma^2)$ and $Cov\ (u_t, u_s) = 0\ \forall\ t \neq s$.

The basic purpose of the Dickey-Fuller test is to verify the null hypothesis of non-stationarity, i.e. the series contains a unit root, against the alternative hypothesis of stationarity, i.e. the root of the characteristic equation lies inside the unit circle:

$$H_0: \phi = 1 \ \ vs \ \ H_1: |\phi| < 1$$

As it is generally easier to test a null hypothesis that a coefficient is equal to zero, y_{t-1} is subtracted from both sides of the equation (2.19) in order to obtain the following regression:

$$y_t - y_{t-1} = \phi y_{t-1} - y_{t-1} + u_t$$
$$\Delta y_t = (\phi - 1) y_{t-1} + u_t = \psi y_{t-1} + u_t$$

in which testing for $\psi = 0$ is equivalent to test for $\phi = 1$ in equation (2.19). Thus, the new hypotheses of the test become:

$$H_0: \psi = 0 \ \ vs \ \ H_1: \psi < 0$$

The test can be conducted allowing for an intercept, or an intercept and a deterministic trend or neither, in the test regression. The general model for the unit root test can be described as follow:

$$y_t = \phi y_{t-1} + \mu + \lambda t + u_t \qquad (2.20)$$

Subtracting y_{t-1} from both sides one obtains:

$$y_t - y_{t-1} = (\phi - 1) y_{t-1} + \mu + \lambda t + u_t$$
$$\Delta y_t = \psi y_{t-1} + \mu + \lambda t + u_t \qquad (2.21)$$

From (2.21) is possible to obtain three model specification, which are applied according to the characteristics of the process considered:

- Unit Root Test (Random Walk): $\Delta y_t = \psi y_{t-1} + u_t$
- Unit Root Test with Drift: $\Delta y_t = \mu + \psi y_{t-1} + u_t$
- Unit Root Test with Drift and Deterministic Time Trend: $\Delta y_t = \mu + \lambda t + \psi y_{t-1} + u_t$

Consider the easiest case in which the estimated regression, that indicates the form in which the regressions is estimated, is described by equation (2.19), while the true process, which describes the null hypothesis under which the distribution is calculated, is a random walk without drift, such that:

Estimated process: $$y_t = \phi y_{t-1} + u_t \qquad (2.22)$$
True process: $$y_t = y_{t-1} + u_t \qquad u_t \sim i.i.d. N(0, \sigma^2) \qquad (2.23)$$

The two main test statistics for the Dickey-Fuller test are the following:

$$\delta_T = T\,(\hat{\phi}_T - 1)$$

$$\tau_T = \frac{(\hat{\phi}_T - 1)}{\hat{\sigma}_{\hat{\phi}_T}}$$

where $\hat{\phi}_T$ is the estimate of the parameter ϕ derived using ordinary least squares (OLS), $\hat{\sigma}_{\hat{\phi}_T}$ is the OLS standard error for the estimated coefficient $\hat{\phi}_T$ and T is the size of the observed series. Consider the OLS estimate of the parameter ϕ:

$$\hat{\phi}_T = \frac{\sum_{t=1}^{T} y_t y_{t-1}}{\sum_{t=1}^{T} y_{t-1}^2}$$

If the true value of ϕ is less than one in absolute value, it is possible to demonstrate that (Hamilton, 1994, p. 216):

$$\sqrt{T}(\hat{\phi}_T - \phi) \xrightarrow{L} N(0, (1 - \phi^2))$$

where $(\hat{\phi}_T - \phi)$ is the deviation of the OLS estimate $\hat{\phi}_T$ from the true value ϕ, T is the size of the observed series and $(\xrightarrow{L})$ is a mathematical symbol which indicates the convergence in distribution (or in law). However, if the true value of ϕ is one, to obtain a nondegenerate asymptotic distribution for $\hat{\phi}_T$, i.e. to obtain a variable with a useful asymptotic distribution, it is necessary to multiply $(\hat{\phi}_T - 1)$ by T rather than by $\sqrt{T}$. This means that the unit root coefficient converges at a faster rate (T) with respect to a coefficient in a stationary regression ($\sqrt{T}$). When the true value of ϕ is one, the difference between the estimate $\hat{\phi}_T$ and the true value can be expressed as (Hamilton, 1994, p. 210):

$$(\hat{\phi}_T - 1) = \frac{\sum_{t=1}^{T} y_{t-1} u_t}{\sum_{t=1}^{T} y_{t-1}^2} \qquad (2.24)$$

from which:

$$T(\hat{\phi}_T - 1) = \frac{T^{-1} \sum_{t=1}^{T} y_{t-1} u_t}{T^{-2} \sum_{t=1}^{T} y_{t-1}^2} \qquad (2.25)$$

At this point, it could be useful to understand why it is necessary scaling equation (2.24) by T rather than $\sqrt{T}$ when the true value of ϕ is one. The process described by equation (2.23) can be rewritten as follow:

$$y_t^2 = (y_{t-1} + u_t)^2 = y_{t-1}^2 + 2u_t y_{t-1} + u_t^2$$

from which:

$$y_{t-1} u_t = \frac{1}{2}(y_t^2 - u_t^2 - y_{t-1}^2) \qquad (2.26)$$

summing equation (2.26) over $t = 1,2,\ldots,T$, one gets:

$$\sum_{t=1}^{T} y_{t-1}u_t = \frac{1}{2}(y_T^2 - y_0^2 - \sum_{t=1}^{T} u_t^2) \tag{2.27}$$

Dividing both sides for T and σ^2, which represents the variance of each random variable u_t, and recalling that $y_0 = 0$, equation (2.27) becomes (Hamilton, 1994, p. 477):

$$\left(\frac{1}{T\sigma^2}\right)\sum_{t=1}^{T} y_{t-1}u_t = \frac{1}{2}\left(\frac{1}{\sqrt{T}\sigma}y_T\right)^2 - \frac{1}{2}\frac{1}{T\sigma^2}\sum_{t=1}^{T} u_t^2 \tag{2.28}$$

where $y_T = (u_T + u_{T-1} + \ldots + u_1) \sim N(0, T\sigma^2)$, implying that:

$$\left(\frac{1}{\sqrt{T}\sigma}y_T\right) \sim N(0,1)$$

so that its square is a chi-square with one degree of freedom (see Hamilton, 1994, p. 746), and so:

$$\left(\frac{1}{\sqrt{T}\sigma}y_T\right)^2 \sim \chi^2(1) \tag{2.29}$$

Consider the second term on the right hand side of the equation (2.28), this term is the sum of independent and identically distributed random variables, as stated in equation (2.23), each one with mean equal to σ^2. From the law of large numbers (see Hamilton, 1994, p. 183) is possible to state that:

$$\frac{1}{T}\sum_{t=1}^{T} u_t^2 \xrightarrow{p} \sigma^2$$

so that

$$\frac{1}{T\sigma^2}\sum_{t=1}^{T} u_t^2 \xrightarrow{p} 1 \tag{2.30}$$

where ($\xrightarrow{p}$) is a mathematical symbol indicating the convergence in probability. Based on (2.29) and (2.30), equation (2.18) becomes:

$$\left(\frac{1}{T\sigma^2}\right)\sum_{t=1}^{T} y_{t-1}u_t \xrightarrow{L} \frac{1}{2}(X - 1) \tag{2.31}$$

where X is a $\chi^2(1)$ random variable.

Finally, from equation (2.25) consider:

$$\sum_{t=1}^{T} y_{t-1}^2$$

Recall that $y_T = (u_T + \ldots + u_1) \sim N(0, T\sigma^2)$, so that $y_{t-1} = (u_{t-1} + \ldots + u_1) \sim N(0, (t-1)\sigma^2)$, from which the expected value of y_{t-1}^2, is $E[y_{t-1}^2] = (t-1)\sigma^2$ which implies that:

$$E\left[\sum_{t=1}^{T} y_{t-1}^2\right] = \sigma^2\,(T-1)\frac{T}{2} = \sigma^2\frac{T^2}{2} - \sigma^2\frac{T}{2} \tag{2.32}$$

Notice that the leading term (see Hamilton. 1994, p.456) in the previous equation is $T^2/2$, thus in order to obtain a random variable with a convergent distribution it is necessary to divide equation (2.32) by T^2, from which one obtains exactly the denominator of the equation (2.25).

Phillips (1987) exploiting the *functional central limit theorem* (see Hamilton, 1994, pp. 479-482) and the *continuous mapping theorem* (see Hamilton, 1994, pp. 482-483) shows that the asymptotic distribution of statistics constructed from unit root processes can be calculated in terms of functionals on standard Brownian motion[3]. Thus, recalling the equation (2.25) it is possible to state that:

$$\left(\frac{1}{T}\right)\sum_{t=1}^{T} y_{t-1}u_t \xrightarrow{L} \frac{1}{2}\sigma^2\{[W(1)]^2 - 1\} \tag{2.33}$$

$$\left(\frac{1}{T}\right)^2\sum_{t=1}^{T} y_{t-1}^2 \xrightarrow{L} \sigma^2\int_0^1 [W(r)]^2 dr \tag{2.34}$$

Since (2.25) is a continuous function of (2.33) and (2.34) it follows that under the null hypothesis of non-stationarity, the OLS estimate of the parameter ϕ is described by (Hamilton, 1994, p. 488):

$$T(\hat{\phi}_T - 1) \xrightarrow{L} \frac{(1/2)\,\{[W(1)]^2 - 1\}}{\int_0^1 [W(r)]^2 dr}$$

The other statistic for testing the null hypothesis that $\phi = 1$ is based on the traditional OLS *t* test of this hypothesis:

$$\tau_T = \frac{(\hat{\phi}_T - 1)}{\hat{\sigma}_{\hat{\phi}_T}} \tag{2.35}$$

Where $\hat{\sigma}_{\hat{\phi}_T}$ is the standard error of the estimated coefficient $\hat{\phi}$, and can be expressed as (see Hamilton, 1994, pp. 488-489):

[3] Standard Brownian motion $W(\cdot)$ is a continuous-time stochastic process, associating each date $r \in [0,1]$ with the scalar $W(r)$ such that: (1) $W(0)=0$; (2) for any dates $0 \le t_1 < t_2 < \ldots < t_k \le 1$ the changes $[W(t_2) - W(t_1)], \ldots, [W(t_{k-1}) - W(t_k)]$ are independent multivariate Gaussian with $[W(s) - W(t)] \sim N(0, s-t)$; (3) for any given realization, $W(t)$ is continuous in t with probability 1 (see Hamilton, 1994, pp. 477-479).

$$\hat{\sigma}_{\hat{\phi}_T} = \sqrt{\frac{s_T^2}{\sum_{t=1}^{T} y_{t-1}^2}} = \sqrt{\frac{\sum_{t=1}^{T}[(y_t - \hat{\phi}_T y_{t-1})^2/(T-1)]}{\sum_{t=1}^{T} y_{t-1}^2}} \tag{2.36}$$

with s_T^2 denoting the OLS estimate of the residual variance. Even if the τ_T *test* is calculated in the same way as the usual *t test*, its limiting distribution is not Gaussian under the null hypothesis of non-stationarity (i.e. when $\phi = 1$). The τ_T *test* described in equation (2.35) is closely related to the δ_T *test* described in equation (2.25), indeed the τ_T *test* is simply the δ_T *test* divided by *T* and by the standard error of the estimated coefficient $\hat{\phi}$ ($\hat{\sigma}_{\hat{\phi}_T}$), that is:

$$\tau_T = \frac{\delta_T}{T\,\hat{\sigma}_{\hat{\phi}_T}}$$

Substituting equation (2.36) one obtains:

$$\tau_T = \frac{T\,(\hat{\phi}_T - 1)}{T}\left\{\frac{s_T^2}{\sum_{t=1}^{T} y_{t-1}^2}\right\}^{-1/2} = T\,(\hat{\phi}_T - 1)\ \frac{\{T^{-2}\sum_{t=1}^{T} y_{t-1}^2\}^{1/2}}{(s_T^2)^{1/2}}$$

Finally, substituting equation (2.25) the result is:

$$\tau_T = \frac{T^{-1}\sum_{t=1}^{T} y_{t-1}u_t}{T^{-2}\sum_{t=1}^{T} y_{t-1}^2}\ \frac{\{T^{-2}\sum_{t=1}^{T} y_{t-1}^2\}^{1/2}}{(s_T^2)^{1/2}} = \frac{T^{-1}\sum_{t=1}^{T} y_{t-1}u_t}{\left\{T^{-2}\sum_{t=1}^{T} y_{t-1}^2\right\}^{1/2}(s_T^2)^{1/2}}$$

where the consistency of $\hat{\phi}_T$ implies that $s_T^2 \xrightarrow{L} \sigma^2$ (see Hamilton, 1994, p. 211).

As in the case of the δ_T *test*, it is possible to write the asymptotic distribution of the τ_T *test* in terms of functional on standard Brownian motion (Hamilton, 1994, p. 489):

$$\frac{(\hat{\phi}_T - 1)}{\hat{\sigma}_{\hat{\phi}_T}} \xrightarrow{L} \frac{(1/2)\,\{[W(1)]^2 - 1\}}{\{\int_0^1 [W(r)]^2 dr\}^{1/2}}$$

In practice, the critical values for statistics that have unknown distribution or a distribution that cannot be expressed in a closed form, as the test statistic $T(\hat{\phi}_T - 1)$ and $\frac{(\hat{\phi}_T - 1)}{\hat{\sigma}_{\hat{\phi}_T}}$, are found using simulation techniques, such as the *Monte Carlo technique*. When some properties of a particular estimation method are unknown or cannot be described analytically, *Monte Carlo technique* is used to create a large enough random sample from the unknow distribution considered in order to simulate these properties. The main steps of the *Monte Carlo technique* are the following (Kočenda and Černý, 2015, p. 193):

1. Random generation of the data with the desired properties using proper data generating process.

2. Performance of the regression and computation of the investigated test statistics (or other parameters of interest).
3. Repetition of the whole procedure, described in the point (1) and (2), *N* times (number of replications). The larger is *N*, the more representative will be the sampled distribution obtained from the values of the test statistics collected. The central idea of the *Monte Carlo techniques* is that of random sampling from a given distribution. Thus, if the number of replications is set too small the final results will be unreliable.

Therefore, it could be of interest to understand how the critical values for the simplest version of the DF *test* analysed in this section (equations (2.22) and (2.23)) are obtained. The first step concerns the generation of a series $\{y_t\}_{t=1}^{T}$ of length T (required number of observations) which follows a unit root process, and such that the series of randomly generated errors $\{u_t\}_{t=1}^{T} \sim N(0, \sigma^2)$. Assuming for simplicity a first value of the series $y_t = 0$ it is possible to construct the series for $\{y_t\}_{t=1}^{T}$ recursively ($y_1 = y_0 + u_1; y_2 = y_1 + u_2; \ldots; y_T = y_{T-1} + u_T$). The second step concerns the estimation of the parameters of the artificial time series $\{y_t\}_{t=1}^{T}$ using OLS and the computation of the *t-statistics* of interest. By performing *N* times this procedure an estimate of the exact small sample distribution of the OLS estimates can be obtained (Hamilton, 1994, pp. 216-217).

It is important to notice that the asymptotic properties of the OLS estimate $\hat{\phi}_T$ when $\phi = 1$ depend on the assumptions that are made about the true model and on the particular specification that is selected in order to estimate the parameter of interest $\hat{\phi}_T$. There are four different cases that can be considered:

- *Case 1*: The estimated process does not contain a constant or a time trend and the true process is a random walk. This is the case analysed in this section.
- *Case 2*: The estimated process contains a constant but not a time trend and the true process is a random walk (see Hamilton, 1994, pp. 490-494).
- *Case 3*: The estimated process contains a constant but not a time trend and the true process is a random walk with drift (see Hamilton, 1994, pp. 495-497).
- *Case 4*: The estimated process contains a constant and a time trend and the true process is a random walk with drift (see Hamilton, 1994, pp. 497-500).

A full set of Dickey-Fuller critical values for the cases reported above (with the exception of the *Case 3*) and considering various sample sizes *T*, obtained using *Monte Carlo technique*, for the *test statistics* $T(\hat{\phi}_T - 1)$ and $(\hat{\phi}_T - 1)/\hat{\sigma}_{\hat{\phi}_T}$ is reported in the *Appendix A* (*Table A1* and *Table A2*). For *Case 3* the estimated coefficients are asymptotically Gaussian, meaning that the standard OLS *t* statistic can be calculated in the usual way and compared with the critical values for the standard *t* distribution.

Based on the critical values obtained with the *Monte Carlo technique* is possible to determine for the two *test statistics* when the null hypothesis of non-stationarity ($\phi = 1$) is accepted or when it is rejected in favour of the alternative of stationarity ($\phi < 1$). For both the *test statistics* considered, if the value of the test is lower (more negative) than the corresponding critical value, at a given level of significance, then the null hypothesis is rejected in favour of the alternative hypothesis. Conversely, if the value of the test is greater (less negative) than the corresponding critical value, at a given level of significance, then the null hypothesis of non-stationarity cannot be rejected, i.e. the null hypothesis of non-stationarity is accepted. For example, under the *Case 1*, assuming a sample size *T*=100, the critical value at the 5% level for the *test statistics* δ_T and τ_T is respectively -7.9 and -1.95. This means that in order to reject the null hypothesis of non-stationarity for a sample of this size the test statistics must be respectively $\delta_T < -7.9$ and $\tau_T < -1.95$.

2.7.2 Augmented Dickey-Fuller test

An important assumption on which is based the ordinary *Dickey-Fuller test* is that the error terms must be uncorrelated, as stated in equation (2.23). However, in practice, it is possible that these error terms in the *DF test* show evidence of serial correlation, leading to unreliable results. The solution proposed by Said and Dickey (1984) is to 'augment' the test including in the regression p lagged differences, such as $\Delta y_{t-1}, \Delta y_{t-2}, \ldots, \Delta y_{t-p}$, in order to ensure that the error term is effectively a white noise process. The resulting test is the so called *Augment Dickey-Fuller test* (ADF).

In order to understand the derivation of the general regression equation used in the *ADF test*, it could be useful to begin with an example. Consider the following AR (3) process:

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \phi_3 y_{t-3} + u_t$$

Subtracting y_{t-1} from both sides of the equation, one gets:

$$y_t - y_{t-1} = (\phi_1 - 1) y_{t-1} + \phi_2 y_{t-2} + \phi_3 y_{t-3} + u_t$$

Add and subtract $\phi_3 y_{t-2}$ to the right hand side of the equation:

$$\Delta y_t = (\phi_1 - 1) y_{t-1} + \phi_2 y_{t-2} + \phi_3 y_{t-3} + \phi_3 y_{t-2} - \phi_3 y_{t-2} + u_t$$
$$\Delta y_t = (\phi_1 - 1) y_{t-1} + (\phi_2 + \phi_3) y_{t-2} - \phi_3 (y_{t-2} - y_{t-3}) + u_t$$

Finally, add and subtract $(\phi_2 + \phi_3)y_{t-1}$:

$$\Delta y_t = (\phi_1 - 1)y_{t-1} + (\phi_2 + \phi_3)y_{t-2} - \phi_3(y_{t-2} - y_{t-3}) + (\phi_2 + \phi_3)y_{t-1} - (\phi_2 + \phi_3)y_{t-1} + u_t$$

$$\Delta y_t = (\phi_1 + \phi_2 + \phi_3 - 1)y_{t-1} - (\phi_2 + \phi_3)(y_{t-1} - y_{t-2}) - \phi_3(y_{t-2} - y_{t-3}) + u_t$$

$$\Delta y_t = (\phi_1 + \phi_2 + \phi_3 - 1)y_{t-1} - (\phi_2 + \phi_3)\Delta y_{t-1} - \phi_3 \Delta y_{t-2} + u_t$$

$$\Delta y_t = \psi y_{t-1} + \sum_{j=1}^{3-1} \gamma_j \Delta y_{t-j} + u_t \qquad (2.37)$$

where $\psi = (\sum_{i=1}^{3} \phi_i) - 1$ and $\gamma_j = -\sum_{k=j+1}^{3} \phi_k$.

The procedure can be generalized to the testing of a single unit root in an *AR* (*p*) process described by the following equation:

$$y_t = \mu + \beta t + \phi_1 y_{t-1} + \ldots + \phi_p y_{t-p} + u_t$$

which becomes:

$$\Delta y_t = \mu + \beta t + \psi y_{t-1} + \sum_{j=1}^{p-1} \gamma_j \Delta y_{t-j} + u_t \qquad (2.38)$$

with

$$\psi = (\sum_{i=1}^{p} \phi_i) - 1 \quad \text{and} \quad \gamma_j = -\sum_{k=j+1}^{p} \phi_k$$

where *p* is the number of autoregressive lag terms incorporated in the test.

From (2.38) is possible to obtain three model specifications for the ADF test, which are applied according to the characteristics of the process considered:

- Unit Root Test (Random Walk): $\Delta y_t = \psi y_{t-1} + \sum_{j=1}^{p-1} \gamma_j \Delta y_{t-j} + u_t$, which is obtained by removing the constant term (μ) and the time trend (βt).
- Unit Root Test with Drift: $\Delta y_t = \mu + \psi y_{t-1} + \sum_{j=1}^{p-1} \gamma_j \Delta y_{t-j} + u_t$, which is obtained by removing the time trend (βt).
- Unit Root Test with Drift and Deterministic Time Trend: $\Delta y_t = \mu + \beta t + \psi y_{t-1} + \sum_{j=1}^{p-1} \gamma_j \Delta y_{t-j} + u_t$.

The basic objective of this test, just like the ordinary DF test, is to evaluate the null hypothesis of non-stationarity against the alternative hypothesis of stationarity in the previous three cases:

$$H_0: \psi = 0 \;\; vs \;\; H_1: \psi < 0$$

The coefficients in equation (2.38) can be consistently estimated by ordinary least squares, and the estimate of the coefficient for y_{t-1} provides a means for testing the null hypothesis of non-stationarity, that is $\psi = (\sum_{i=1}^{p} \phi_i) - 1 = 0$. The resulting *t-type statistic* has the same asymptotic critical values of the corresponding ordinary *DF test* (see *Appendix A*, *Table A2*), and it can be computed as:

$$\tau = \frac{\hat{\psi}}{\hat{\sigma}_{\hat{\psi}}}$$

where $\hat{\psi}$ is the OLS estimate of the parameter ψ and $\hat{\sigma}_{\hat{\psi}}$ is the standard error of the estimated coefficient $\hat{\psi}$. The null hypothesis of an integrated process is rejected in favour of the alternative hypothesis of stationarity whenever the test statistics is lower than the critical value at a given level of significance. Conversely, if the test statistic is greater than the critical value at a given level of significance, the null hypothesis cannot be rejected.

The main problem to face when utilizing the ADF test is how to decide the optimal lag length of the dependent variable. Two rules of thumb are suggested by Brooks (2008, p. 329) to overcome this problem:

- First, the frequency of the data can be used to determine the adequate number of lags to use. For example, if the data are monthly use 12 lags, if the data are quarterly use 4 lags and so on.
- Second, the optimal number of lags can be found choosing the number of lags that minimizes the value of an information criterion.

As argued by Verbeek (2017, p. 305) the choice of the optimal lag length is extremely important, if too many lags are considered this will reduce the power of the test, but, if too few lags are included the asymptotic distributions from the table in the *Appendix A* (*Table A1* and *Table A2*) are not valid, and the test may lead to seriously biased results.

2.7.3 Phillips-Perron test

An alternative to the *Augmented Dickey-Fuller tests* is the *Phillips-Perron test* (PP) named after the two authors who first proposed it in 1988. The main difference between the PP unit root tests and the ADF tests is the treatment of serial correlation and heteroskedasticity in the error terms. Instead of adding additional lagged differences in the regression in order to obtain a white noise error term, Phillips and Perron (1988), starting from the same regression considered for the Dickey-Fuller test, make a non-parametric correction of the Dickey-Fuller test statistics to take into account the potential autocorrelation pattern in the errors (Verbeek, 2017, p. 306).

To illustrate the idea of the *Phillips-Perron test* consider the following regression:

$$y_t = \mu + \phi y_{t-1} + u_t \qquad (2.39)$$

where μ is a constant, ϕ is the coefficient of the term y_{t-1} and u_t may be serially correlated and possibly heteroskedastic. The assumption regarding the true process are that $\mu = 0$ and $\phi = 1$, meaning that the true process can be described as:

$$y_t = y_{t-1} + u_t$$

In other words, this specification considers the case in which the estimated process contains a constant but not a time trend and the true process is a random walk (see also *Case 2* in Section 2.7.1). Phillips and Perron (1988) suggest estimating equation (2.39) by OLS even if u_t are serially correlated and then directly modifying the statistics (τ_T *test* and δ_T) to account for the serial correlation. For completeness these modified statistics, denoted respectively Z_τ and Z_π, are reported below, but they will not be explained in detail since this is beyond the objectives of the thesis (see Hamilton, 1994, pp. 509-510):

$$Z_\tau = \left(\frac{\gamma_0}{\lambda^2}\right)^{1/2} \tau_T - \frac{1}{2}\left(\frac{\lambda^2 - \gamma_0}{\lambda}\right)\left(\frac{T\,\hat{\sigma}_{\hat{\phi}_T}}{s_T}\right) \qquad (2.40)$$

$$Z_\pi = T(\hat{\phi}_T - 1) - \frac{1}{2}\left(\frac{T^2\,\hat{\sigma}_{\hat{\phi}_T}}{s_T^2}\right)(\lambda^2 - \gamma_0) \qquad (2.41)$$

where τ_T is the DF *t-type statistic*, s_T^2 is the OLS estimate of the variance of u_t, $\hat{\sigma}_{\hat{\phi}_T}$ is the OLS standard error for the estimated coefficient $\hat{\phi}_T$, T is the size of the observed series, $\gamma_0 = E(u_t^2)$, λ^2 is the asymptotic variance of the sample mean ($\bar{u}$) of *u*, that is (Hamilton, 1994, p. 510):

$$\sqrt{T}\bar{u} = T^{-1/2}\sum_{t=1}^{T} u_t \xrightarrow{L} N(0, \lambda^2)$$

The statistics (2.40) and (2.41) require knowledge of the population parameters γ_0 and λ^2. Even if these parameters are unknown they are easy to estimate, in fact the sample variance of the OLS residual $\hat{u}_t$ is a consistent estimate of γ_0, the Newey-West variance estimate (see Hamilton, 1994, pp. 281-283) of u_t using $\hat{u}_t$ is a consistent estimate of λ^2. The resulting estimates $\hat{\gamma}_0$ and $\hat{\lambda}^2$ can be used in the equations (2.40) and (2.41) to calculate the *test statistics*, which under the null hypothesis that $\mu = 0$ and $\phi = 1$ have the same asymptotic distribution as the corresponding Dickey-Fuller tests (τ_T and δ_T), which in the case considered is the *Case 2* described in Section 2.7.1. If the PP statistics is statistically significant, i.e. if the test statistic considered is lower than the corresponding critical values at a given level of significance, it is possible to reject the null hypothesis of unit root, even in the presence of serial correlation and/or heteroskedasticity (Kočenda and Černý, 2015, p. 72).

There are two important advantages that the PP tests has over the ADF tests: first, the PP tests are robust to general forms of heteroskedasticity in the error term; second, in the PP tests there is no need to define the optimal lag length for the regression (Zivot and Wang, 2006, p. 123).

Despite the advantages, the ADF test is generally preferred to the PP test because there is good evidence that the *Phillips-Perron tests* perform less well in finite sample than the *Augmented Dickey-Fuller tests* (Davidson and Mackinnon, 2004, p. 613).

2.8 Stationarity Tests

The technical approach of the stationarity tests is completely different from that of the unit root tests, the main difference is the transposition of the null and the alternative hypothesis. While the unit root tests evaluate the null hypothesis that a time series is non-stationary against the alternative hypothesis of stationarity, the stationarity tests evaluate the null hypothesis that the time series is an integrated process of order zero, i.e. stationary, against the alternative hypothesis of a unit root.

The most commonly used stationarity test is the *KPSS test* that owes its name to Kwiatkowski, Phillips, Schmidt and Shin (1992). This approach has been developed in order to overcome the low test power of the unit root tests, a time series with a root close to one that was typically found non-stationary with the ADF test and the PP test can be correctly found stationary with the KPSS test (Kočenda and Černý, 2015, p. 73).

The *KPSS test* is derived by the following model (Kwiatkowski *et al*, 1992):

$$y_t = \xi t + r_t + \varepsilon_t$$
$$r_t = r_{t-1} + u_t \quad \text{with } u_t \text{ are } \textit{i.i.d. } N(0, \sigma_u^2)$$

where y_t, with $t = 1,2,\ldots,T$, is the observed series that has to be test for stationarity. This series is decomposed into the sum of a deterministic trend (ξt), a random walk (r_t) and the error term (ε_t) which may be serially correlated and possibly heteroskedastic, as for the *Phillips-Perron test* (see Section 2.7.3). The initial value of r_t, that is r_0, is treated as fixed and serves the role of an intercept.

The null hypothesis of stationarity (or trend stationarity) of the time series $\{y_t\}_{t=-\infty}^{\infty}$ corresponds to the hypothesis that the variance of the random walk $\{r_t\}_{t=-\infty}^{\infty}$ equals zero, which implies that $\{r_t\}_{t=-\infty}^{\infty}$ is a constant:

$$H_0: \sigma_u^2 = 0 \;\; vs \;\; H_1: \sigma_u^2 > 0$$

The test statistic is given by (Kwiatkowski *et al.*, 1992):

$$KPPS = T^{-2} \sum_{t=1}^{T} \frac{S_t^2}{\hat{\sigma}^2}$$

where $S_t = \sum_{s=1}^{t} \hat{\varepsilon}_s$ is the partial sum process of the OLS residuals ($\hat{\varepsilon}_t$) from the regression of $\{y_t\}$ on an intercept and time trend. The authors define the 'long-run variance' (σ^2) as:

$$\sigma^2 = \lim_{T\to\infty} E(S_T^2)$$

so that $\hat{\sigma}^2$ is a consistent estimator of σ^2 which can be constructed from the residuals ($\hat{\varepsilon}_t$), as in Phillips and Perron (1988) (see Section 2.7.3).

The asymptotic distribution of the *KPSS test* is nonstandard, and Kwiatkowski *et al.* (1992) calculate the critical values via direct simulation, using sample size of 2000 and 50000 replications. These critical values are reported in *Appendix A* (*Table A3*).

Although the *KPSS test* resolves the low test power of the unit root tests, it is important to be aware that any results of statistical testing are just probabilistic, meaning that there is always a non-zero chance of being wrong. A clever approach that should be used in unit root testing is to combine the *ADF test* (or the *PP test*) and the *KPSS test*. If a time series is found stationary with the *ADF test* then it will be more likely to find stationarity also using the *KPSS test*. Similarly, if a time series is found non-stationary using the *KPSS test* then it is reasonable to expect non-stationarity also with the *ADF test*. Nonetheless, it can happen that a time series that was found stationary with the *ADF test* will be marked as non-stationary using the *KPSS test*. In such cases it is important to be very careful with the final conclusions (Kočenda and Černý, 2015, p. 73).

2.9 Spurious Regressions

The usual statistical results for the linear regression model are based on the assumption that the variables x_t and y_t are stationary, that is $x_t \sim I(0)$ and $y_t \sim I(0)$. Consider a regression in the following form:

$$y_t = x_t'\beta + u_t$$

where the elements of y_t and x_t might be nonstationary. If there does not exist some value for β for which the residual is stationary, i.e. $u_t = y_t - x_t'\beta \ \sim I(0)$, then the OLS estimator is prone to generate *spurious regression* (Hamilton, 1994, p. 557).

Consider the variables x_t and y_t, generated by two independent random walks:

$$y_t = y_{t-1} + u_{1t} \qquad u_{1t} \sim IID\ (0, \sigma_1^2)$$
$$x_t = x_{t-1} + u_{2t} \qquad u_{2t} \sim IID\ (0, \sigma_1^2)$$

where ε_{1t} and ε_{2t} are mutually independent disturbance terms. The process generating these two variables are independent and so there is nothing that leads to a relationship between them. Suppose that one of those variables is regressed on the other such that:

$$y_t = \alpha + \beta x_t + u_t$$

One should expect to find no evidence of a relationship, so that the estimate of β is near to zero and its associated t-statistics is insignificant. However, as argued by Granger and Newbold (1974), these types of regression, which relate nonstationary time series, frequently have high R^2 and also highly autocorrelated residuals (indicated by a very low Durbin-Watson statistics). In such situations the usual significance tests about regression coefficients can be very misleading. Granger and Newbold (1974) conducting sampling experiments demonstrate that the traditional significance tests are severely bias towards rejection of the null hypothesis of no relationship, i.e. acceptance of a spurious relationship. According to Phillips (1986), the reason for these ambiguous results is that the distributions of the conventional test statistic under the assumption of no-stationarity are very different from those derived under the assumption of stationarity. In particular, the following results about the behaviour of the OLS estimator are due to Phillips (1986):

- The OLS estimator of the coefficients do not converge in probability to constants as $T \to \infty$.
- The conventional OLS t-statistics used to assess the significance of the coefficients in the regression analysis do not have well defined asymptotic distributions.
- Low value for the Durbin Watson statistics and relatively high values coefficient of determination R^2 are expected in spurious regressions.

Because x_t and y_t contain stochastic trend, the OLS estimator tends to find a significant correlation between them, even if they are totally unrelated. This means that the final results seem to be good under standard measures, but they are valueless (Verbeek, 2017, p. 352). Statistically, the problem is that u_t is nonstationary because both x_t and y_t are nonstationary process. According to Hamilton (1994, pp. 561-562) one approach to avoid the problems of spurious regressions is to include lagged values of both the dependent and the independent variables:

$$y_t = \alpha + \phi y_{t-1} + \beta x_t + \delta x_{t-1} + u_t$$

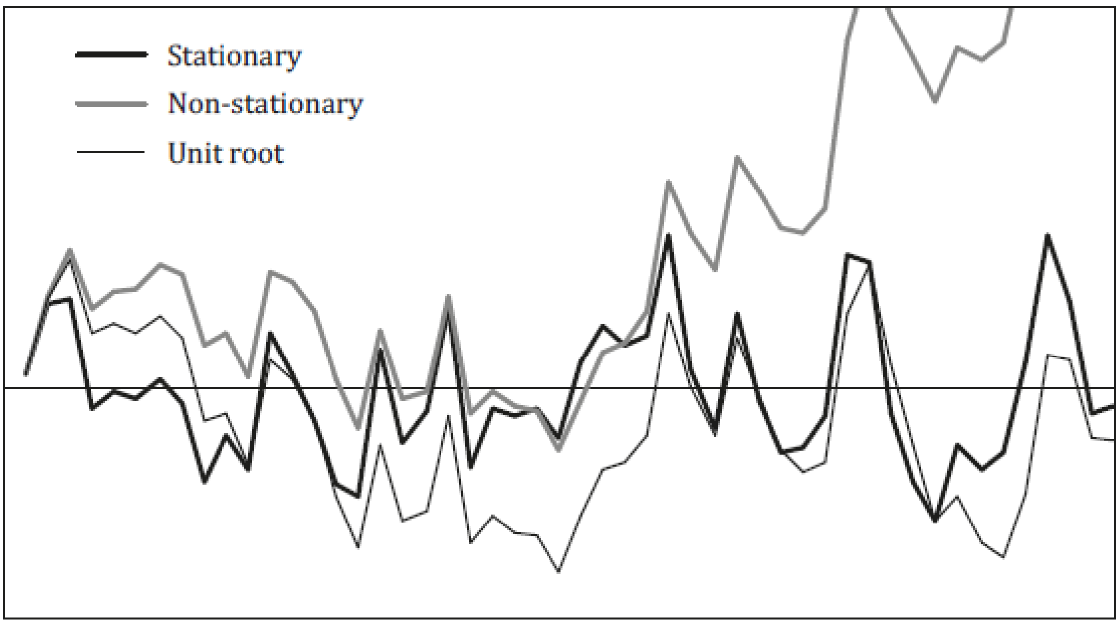
In this case, there are no spurious regression problem because there exist coefficients ($\phi = 1$, $\beta = \delta = 0$) for which $u_t \sim I(0)$. Thus, the OLS estimator is consistent for all parameters.

2.10 Cointegration

2.10.1 Definition of Cointegration

A peculiar feature of most macroeconomic time series is the presence of a unit root in levels, that is $k_t \sim I(1)$, and stationarity in first differences, that is $\Delta k_t \sim I(0)$. This kind of time series has very unpleasant characteristics; consider a time series $k_t \sim I(1)$ with $k_0 = 0$: it does not exhibit mean reversion even if it could have a constant long term mean (e.g. a random walk with drift), its variance goes to infinity as *t* goes to infinity, an innovation has a permanent effect on the value of the time series (see Section 2.6), the expected time between two consecutive crossings of the value $k_0 = 0$ is infinity and finally the autocorrelations ρ_s are not independent of time and converge to one for all *s* as $t \to \infty$ (Engle and Granger, 1987). These unappealing features are the reason why nonstationary time series need to be manipulated (*differencing* or *detrending* (see Section 2.6)) in order to reduce them to stationary time series which can be analysed more easily. If a time series $\{w_t\}_{t=-\infty}^{\infty}$ is weakly stationary its variance is finite, an innovation has only a temporary effect on its value, this means that any shock that occurs at time *t*, intended as a deviation from the *long-run mean* $E(y_t) = \mu$, has a diminishing effect over time and finally disappears at time *t*+*s* as $s \to \infty$, bringing the time series back to its long term equilibrium, i.e. stationary time series has limited memory of its past behaviour (see Section 2.6) and the expected time between two consecutive crossings of the long-run mean is finite (Engle and Granger, 1987). This property of stationary time series to return to its long-run mean is called *mean reversion.*

Figure 2.1: *Stationarity, Non-Stationarity and Unit Root Time Series: A Comparison*



Source: *E. Kočenda, A. Černý (2015)*)

In order to better understand the property of mean reversion, it can be useful to consider the example illustrated in *Figure 2.1*. Figure 2.1 shows examples of stationary and nonstationary (*Non-stationarity* and *Unit root*) time series. The line labelled as '*Non-stationary'* referrer to a random walk in the explosive case scenario (see Section 2.6), while the line labelled as '*Unit root*' referred to an integrated process of order one (see Section 2.5). The stationary time series tends to return often to its initial value (i.e. the long-run mean), while the '*Non-stationary*' time series explodes after a very short period of time. Finally, the time series containing a unit root seems to have a trend very similar to the stationary time series, but it does not return to its initial value as often. Being on the edge between stationary and nonstationary time series, unit root processes play a crucial role in time series analysis and this is the reason why testing for unit root is one of the most important tasks for this type of analysis (see Section 2.7).

Since integrated of order one processes have infinite variance, while stationary processes have finite variance, it is always true that if $w_t \sim I(0)$ and $k_t \sim I(1)$, their sum will be an integrated process of order one (Engle and Granger, 1987).

If $\{w_t\}_{t=-\infty}^{\infty}$ and $\{k_t\}_{t=-\infty}^{\infty}$ are both integrated processes of order *d*, i.e. $I(d)$, then it is generally true that the linear combination:

$$g_t = w_t - ak_t$$

will also be an integrated process of order *d*. However, in some cases, it is possible that some linear combinations of $\{w_t\}_{t=-\infty}^{\infty}$ and $\{k_t\}_{t=-\infty}^{\infty}$ produce a time series $\{g_t\}_{t=-\infty}^{\infty}$ that can be $g_t \sim I(d-b)$ with $b > 0$.

If such a linear combination exists, then the time series are called *cointegrated*. The concept of *cointegration* was first introduced by Granger (1981) and Granger and Weiss (1983), but the classical reference is the seminal paper by Engle and Granger (1987).

Engle and Granger (1987) propose the following definition: *"The components of the vector* x_t *are said to be co-integrated of order d, b, denoted* $x_t \sim CI(d,b)$*, if* (i) *all the components of* x_t *are I(d);* (ii) *there exists a vector* $\alpha(\neq 0)$ *so that* $z_t = \alpha' x_t \sim I(d-b)$*, b>0. The vector* α *is called the co-integrated vector."* In other words, the N components of a vector $x_t = (x_{1t}, x_{2t}, \ldots, x_{Nt})$ are said to be cointegrated of order (*d*, *b*), denoted $CI(d,b)$, if each element of x_t is *I(d)*, and if a linear combination $z_t = \beta_1 x_{1t} + \beta_2 x_{2t} + \ldots + \beta_N x_{Nt}$ exists that is $I(d-b)$. The vector $\beta = (\beta_{1t}, \beta_{2t}, \ldots, \beta_{Nt}) \neq 0$ is a *N*x*1* cointegrating vector (Kočenda and Černý, 2015, p. 157).

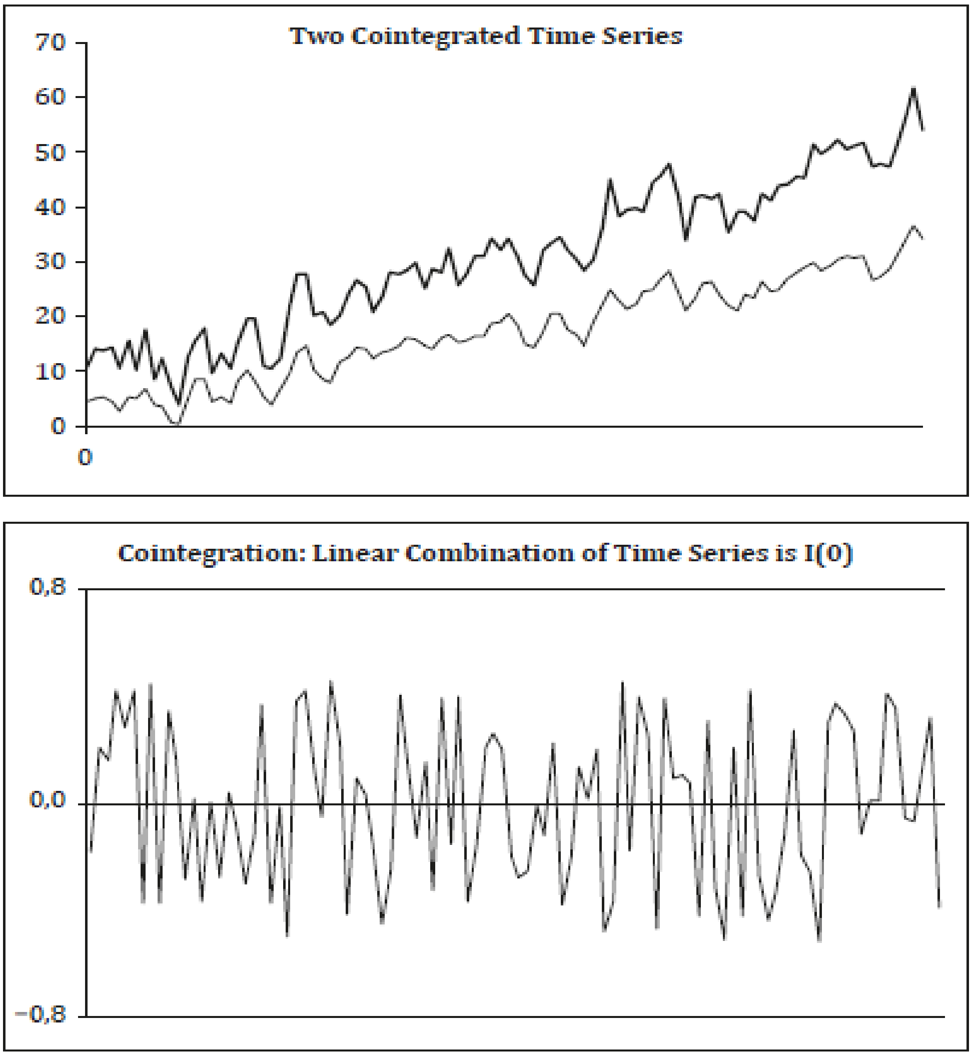
Since the generalization to orders of integration higher than one does not have remarkable economics applications, because integrations of order two or higher are very rare in economics, from now on it will be analysed the case of integrated time series of order one.

Consider two time series $\{x_t\}_{t=-\infty}^{\infty}$ and $\{y_t\}_{t=-\infty}^{\infty}$ both $I(1)$, they are said to be cointegrated of order (1, 1), denoted $CI(1,1)$, if there exists a non-zero linear combination $z_t = \beta_1 x_t - \beta_2 y_t$ that is stationary, where the vector $\beta = (\beta_1, \beta_2)$ is the cointegrating vector. Intuitively, the two time series individually are nonstationary but 'move together' over time, in the sense that there exist some influences on them, which imply that they do not drift too far away from each other except for transitory fluctuations. The cointegrating linear combination defines a sort of long run equilibrium relationship to which the cointegrating variable will be forced to return, despite short run deviations (Brooks, 2008, p. 336).

Consider the cointegrating linear combination:

$$z_t = (\beta_1 x_t - \beta_2 y_t) \sim I(0)$$

Figure 2.2: *Examples of cointegrated time series*



Source: *E. Kočenda, A. Černý (2015)*

This is a time series of the short-run deviations of $\{x_t\}$ and $\{y_t\}$ from their long-run equilibrium which is defined by $(\beta_1 x_t - \beta_2 y_t) = \alpha$, where α is a constant which can assume any real value. Since $\{z_t\}_{t=-\infty}^{\infty}$ is a stationary process it can deviate in the short-run from its equilibrium value (α), but in the long-run it tends to return to it (Kočenda and Černý, 2015, p. 158).

Figure 2.2 provides an intuitive illustration of the property of cointegration. The first panel plots two upward trending time series, it is easy to notice that these processes are nonstationary since they wander far from their respective starting value but they do not drift too far away from each other; the second panel plots the stationary linear combination of the two time series which exhibits the typical mean reverting behaviour around the long-term equilibrium.

2.10.2 Properties of the Cointegrating Vector

Consider the (*N*x*1*) vector $x_t = (x_{1t}, x_{2t}, \dots, x_{Nt}) \sim I(1)$. The vector x_t is cointegrated if there exist an (*N*x*1*) vector $\beta = (\beta_1, \beta_2, \dots, \beta_N)' \neq 0$ such that:

$$\beta' x_t = \beta_1 x_{1t} + \beta_2 x_{2t}, \dots + \beta_N x_{Nt} \sim I(0) \qquad (2.42)$$

The cointegrating vector β is not unique, if $\beta' x_t$ is stationary then so is $c\beta' x_t$ for any nonzero scalar *c*, this means that if β is a cointegrating vector, then so is $c\beta$. Hence, in order to uniquely identify β, an arbitrary normalization must be made. A typical approach is to normalize the cointegrating vector such that the first element of β is one (Hamilton, 1994, p. 574):

$$\beta = (1, -\beta_2, \dots, -\beta_N)'$$

Then, the equilibrium relationship can be written as:

$$\beta' x_t = x_{1t} - \beta_2 x_{2t}, \dots - \beta_N x_{Nt} \sim I(0)$$

If there are more than two variables contained in x_t there might be more than one cointegrating linear relationship, meaning that there might be more than one equilibrium relationship. In fact, because there are *N* variables in the process $\{x_t\}$, at least in principle, the number on linearly independent cointegrating relationship h (cointegrating rank) can range from 0 to *N-1*.
Consider a generic h, such that $0 < h \leq N - 1$. In this case, the cointegrating vector β becomes an Nxh cointegration matrix:

$$\beta = (\beta_i^{(1)}, \beta_i^{(2)}, \beta_i^{(3)}, \dots, \beta_i^{(h)}) \qquad \text{with } i = 1,2,3, \dots, N$$

Again, the cointegrating vectors and consequently the equilibrium relationships are not uniquely identified (see Section 2.11.2). Any linear combination of the cointegrating vectors is a cointegrating vector as well. Moreover, if $\beta' x_t$ is stationary, then for any nonzero (*1* x *h*) vector *k'*, the scalar $k'\beta' x_t$ is also stationary. Then the (*N* x *1*) vector $\alpha' = k'\beta'$ could also be described as a cointegrating vector (Hamilton, 1994, p. 574).

2.10.3 Common Trends Model

The common trends model is an approach introduced by Stock and Watson (1988) to model the cointegration. The idea behind the common-trends representation is that a time series can be expressed as a sum of two component time series: a stationary component and a nonstationary component. Consider two time series:

$$\begin{aligned} x_t &= \eta_{x_t} + \varepsilon_{x_t} \\ y_t &= \eta_{y_t} + \varepsilon_{y_t} \end{aligned} \qquad (2.43)$$

Where $\{\varepsilon_{x_t}\}$ and $\{\varepsilon_{y_t}\}$ represents the stationary components and $\{\eta_{x_t}\}$ and $\{\eta_{y_t}\}$ represent the random walk components or stochastic trend component (nonstationary components) defined by the equations:

$$\eta_{x_t} = (\eta_{x_{t-1}} + \omega_{x_t}) \sim I(1)$$
$$\eta_{y_t} = (\eta_{y_{t-1}} + \omega_{y_t}) \sim I(1)$$

Suppose that the linear combination $x_t = \gamma y_t$ is the cointegrating combination that results in a stationary process. Substituting the equations (2.43) in the linear combination and rearranging the terms, one obtains:

$$x_t - \gamma y_t = \eta_{x_t} + \varepsilon_{x_t} - \gamma(\eta_{y_t} + \varepsilon_{y_t})$$
$$x_t - \gamma y_t = (\eta_{x_t} - \gamma\eta_{y_t}) + (\varepsilon_{x_t} - \gamma\varepsilon_{y_t})$$

Clearly, the two series cannot be cointegrated unless $\eta_{x_t} = \gamma\eta_{y_t}$, that is the nonstationary component must be zero. This means that the two stochastic trends must be generated by the same random walk processes and can only differ by a linear scaling factor γ (Kočenda and Černý, 2015, p. 159).

$$x_t - \gamma y_t = (\varepsilon_{x_t} - \gamma\varepsilon_{y_t}) \sim I(0)$$

In the common trends model the cointegrating linear composition acts to nullify the nonstationary components (Vidyamurthy, 2004, pp. 78-79).

2.10.4 Error Correction Model

The cointegration dynamics can be explained using the notion of error correction. The idea is that cointegrated processes have a long-run equilibrium, represented by the long-run mean of the linear combination of the processes involved. If a short-run deviation from the long-term equilibrium occurs, then the series involved will adjust themselves in order to restore the equilibrium (Vidyamurthy, 2004, p. 76). The error correction model links the long-run equilibrium with the short-run dynamic adjustment mechanism which explains how the variables involved react when they deviate from the equilibrium.

In a seminal paper Engle and Granger (1987) demonstrate that cointegrated time series can always be represented by an error correction model and that the existence of an error correction model always implies cointegration.

Consider a bivariate vector integrate of order one $y_t = (y_{1t}, y_{2t})'$ which is cointegrated with cointegrating vector $\beta = (1, -\beta_2)'$ so that:

$$\beta y_t = (y_{1t} - \beta_2 y_{2t}) \sim I(0)$$

To derive a simple version of the error correction model, consider the following vector auto-regressive model:

$$y_{1t} = \mu_1 + a_{11}y_{1t-1} + a_{12}y_{2t-1} + \varepsilon_{1t}$$
$$y_{2t} = \mu_2 + a_{21}y_{1t-1} + a_{22}y_{2t-1} + \varepsilon_{2t}$$

Subtracting y_{1t-1} from both sides of the first equation and y_{2t-1} from both sides of the second equation, one obtains:

$$y_{1t} - y_{1t-1} = \mu_1 + (a_{11} - 1)y_{1t-1} + a_{12}y_{2t-1} + \varepsilon_{1t}$$
$$y_{2t} - y_{2t-1} = \mu_2 + (a_{21} - 1)y_{1t-1} + a_{22}y_{2t-1} + \varepsilon_{2t}$$

which can be expressed in matrix form as:

$$\begin{pmatrix} \Delta y_{1t} \\ \Delta y_{2t} \end{pmatrix} = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} + \begin{pmatrix} (a_{11} - 1) & a_{12} \\ a_{21} & (a_{22} - 1) \end{pmatrix} \begin{pmatrix} y_{1t-1} \\ y_{2t-1} \end{pmatrix} + \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix}$$

setting $(a_{11} - 1) = -\frac{a_{12}a_{21}}{1-a_{22}}$, one gets:

$$\Delta y_{1t} = \mu_1 + -\frac{a_{12}a_{21}}{1 - a_{22}} y_{1t-1} + a_{12}y_{2t-1} + \varepsilon_{1t}$$

$$\Delta y_{2t} = \mu_2 + a_{21}y_{1t-1} - (1 - a_{22})y_{2t-1} + \varepsilon_{2t}$$

Assuming $a_{12} \neq 0$ and $a_{21} \neq 0$, it is possible to normalize the cointegrating vector with respect to either variable, obtaining the error correction representation. For example, normalizing with respect to y_{1t-1} the resulting error correction model is:

$$\Delta y_{1t} = \mu_1 + \alpha_1(y_{1t-1} - \beta_2 y_{2t-1}) + \varepsilon_{1t} = \mu_1 + \alpha_1 z_{t-1} + \varepsilon_{1t} \quad (2.44)$$
$$\Delta y_{2t} = \mu_2 + \alpha_2(y_{1t-1} - \beta_2 y_{2t-1}) + \varepsilon_{2t} = \mu_2 + \alpha_2 z_{t-1} + \varepsilon_{2t} \quad (2.45)$$

Where $\beta_2 = \frac{(1-a_{22})}{a_{21}}$, $\alpha_1 = -\frac{a_{12}a_{21}}{1-a_{22}}$ and $\alpha_2 = a_{21}$ are the adjustment coefficients, indicative of the speed at which the time series correct themselves to maintain equilibrium, ε_{1t} and ε_{2t} are white noise disturbances, μ_1 and μ_2 are constants, and $\{z_t\}$ is the equilibrium cointegrating linear combination $(y_{1t} - \beta_2 y_{2t})$, $\{z_t\}$ is a stationary process denoting the deviation from the long-term equilibrium. From this analysis it is easy to see that equations (2.44) and (2.45) indicate that the current changes of y_{1t} and y_{2t}, respectively Δy_{1t} and Δy_{2t}, are proportional to the previous deviation from the equilibrium (Kočenda and Černý, 2015, p. 159).

A significant problem that might arise with the simplest version of the error correction model is that this specification might not be sufficient to assure that ε_{1t} and ε_{2t} are white noise disturbances. A solution to overcome this problem is to include in the model p lags of Δy_{1t} and Δy_{2t} in order to ensure that the disturbances are effectively white noise processes (this is the same approach used to proceed from the ordinary *Dickey-Fuller test* to the *Augmented Dickey-Fuller test*) (Zivot and Wang, 2006, p. 437):

$$\Delta y_{1t} = \mu_1 + \alpha_1(y_{1t-1} - \beta_2 y_{2t-1}) + \sum_{j=1}^{p} \gamma_{11}^{j} \Delta y_{1t-j} + \sum_{j=1}^{p} \gamma_{12}^{j} \Delta y_{2t-j} + \varepsilon_{1t}$$

$$\Delta y_{2t} = \mu_2 + \alpha_2(y_{1t-1} - \beta_2 y_{2t-1}) + \sum_{j=1}^{p} \gamma_{21}^{j} \Delta y_{1t-j} + \sum_{j=1}^{p} \gamma_{22}^{j} \Delta y_{2t-j} + \varepsilon_{2t}$$

where $\Gamma_j = \begin{pmatrix} \gamma_{11}^{j} & \gamma_{12}^{j} \\ \gamma_{21}^{j} & \gamma_{22}^{j} \end{pmatrix}$ are *2*x*2* matrixes of autoregressive coefficients.

Finally, it is also possible to expand the error correction model to the general case with *N* cointegrated variables:

$$\Delta y_t = \mu + \alpha\beta' y_{t-1} + \sum_{j=1}^{p} \Gamma_j \Delta y_{t-j} + \varepsilon_t \tag{2.46}$$

where $y_t = (y_{1t}, y_{2t}, \dots, y_{Nt})'$ represents the vector of the *N* cointegrated variables, $\mu = (\mu_1, \mu_2, \dots, \mu_N)'$ represents the vector of the *N* intercepts, $\varepsilon_t = (\varepsilon_{1t}, \varepsilon_{2t}, \dots, \varepsilon_{Nt})'$ represents the vector of the *N* disturbances, $\beta = (\beta_i^{(1)}, \beta_i^{(2)}, \beta_i^{(3)}, \dots, \beta_i^{(h)})$ with *i=1,2,…,N* represents the $N \times h$ cointegrating matrix, α is a $N \times h$ matrix of the adjustment coefficients and

$$\Gamma_j = \begin{pmatrix} \gamma_{11}^{j} & \cdots & \gamma_{1N}^{j} \\ \vdots & \ddots & \vdots \\ \gamma_{N1}^{j} & \cdots & \gamma_{NN}^{j} \end{pmatrix}$$

are the $N \times N$ matrixes of autoregressive coefficients (Kočenda and Černý, 2015, p. 159).

2.11 Testing for Cointegration

2.11.1 Engle-Granger Methodology

Testing for cointegration means testing for the existence of long-run equilibria among the elements of one or more time series. The testing methodology proposed by Engle and Granger (1987) is a simple residual-based testing method which only consider the case with at most one cointegrating vector. The testing methodology proposed by Engle and Granger (1987) enables to distinguish if two or more variables are cointegrated and it allows for the estimation of a very simple form of error correction representation. The main problem of this technique is that it allows to estimate only one cointegrating vector, meaning that if a system of *N* > 2 cointegrated variables is considered, in which there could be up to *N-1* cointegrating vectors, the Engle-

Granger methodology can only detect the presence of cointegration among the variables but it is unable to test for the number of cointegrating vectors.

Following Kočenda and Černý (2015, pp. 162-165) the Engle-Granger methodology for the simple case of two cointegrated variables, x_t and y_t, can be divided into four steps:

1. The first step is to test the two variables individually for their order of integration. In order to proceed with the cointegration test x_t and y_t should be integrated of order one, because if the individual time series are integrated of different orders then in can be concluded that they are not cointegrated (see Section 2.10.1). To ascertain the order of integration it is possible to use all the unit root tests discussed earlier, e.g. the *Augmented Dickey-Fuller test* (Section 2.7.2) or the *Phillips Perron test* (Section 2.7.3). In general, a good approach is to combine unit root tests and stationarity tests, such as the *KPSS test* (Section 2.8).If the null hypothesis of a unit root is individually accepted for both variables, meaning that $x_t \sim I(1)$ and $y_t \sim I(1)$, it is possible to proceed to the next step.

2. The second step concerns the estimation of the long run equilibrium relationship between the time series using OLS. Two different specifications can be considered:

$$y_t = \mu + \beta x_t + \varepsilon_t \tag{2.47}$$
$$y_t = \mu + \gamma t + \beta x_t + \varepsilon_t \tag{2.48}$$

Equation (2.47) considers only the presence of a constant while equation (2.48) controls for possible linear time trends in the time series. In general, if the two series appear to be trending then the proper specification to be estimated should be equation (2.48), otherwise should be considered equation (2.47).

Engle and Granger (1987) demonstrate that if two variables are cointegrated, then OLS estimator produce super-consistent estimate of the cointegrating vector, which means that the OLS estimate $\hat{\beta}$ converges to the true value β faster than it would if the series were stationary. The parameter estimates in equation (2.47) are estimated from:

$$\hat{\beta} = \frac{\sum(x_t - \bar{x}_t)(y_t - \bar{y}_t)}{\sum(x_t - \bar{x}_t)^2} \qquad \hat{\mu} = \bar{y}_t - \hat{\beta}\,\bar{x}_t$$

where $\bar{x}_t$ and $\bar{y}_t$ is the mean of x_t and y_t respectively. Thus, the estimated regression is given by:

$$\hat{y}_t = \hat{\mu} + \hat{\beta} x_t$$

Finally, the residuals can be estimated as:

$$\widehat{\varepsilon_t} = y_t - \hat{y}_t$$

3. The third step concerns the application of a unit root test, for example the *Augmented Dickey Fuller test*, on the estimated residuals ($\widehat{\varepsilon_t}$) of the OLS regression in order to decide whether the residuals are stationary or not. If the residuals are stationary then x_t and y_t are cointegrated, while if the residuals are nonstationary then x_t and y_t are not cointegrated.

Recall that unit root tests are statistical tests of the null hypothesis that a time series is nonstationary against the alternative that it is stationary. Thus, the hypotheses to be tested are:

$H_0: \widehat{\varepsilon_t} \sim I(1)$ (no stationarity)

$H_1: \widehat{\varepsilon_t} \sim I(0)$ (stationarity)

Under the null hypothesis a stationary linear combination of x_t and y_t has not been found. This means that if H_0 is not rejected, there is no cointegration between the variables. Conversely, if H_0 is rejected this means that a stationary linear combination of x_t and y_t has been found and so the two variables are cointegrated (Brooks, 2008, p. 340).

The *ADF test* is performed on the following model:

$$\Delta\widehat{\varepsilon_t} = \psi\hat{\varepsilon}_{t-1} + \sum_{j=1}^{k} \gamma_j \Delta\hat{\varepsilon}_{t-j} + u_t \tag{2.49}$$

where $\Delta\widehat{\varepsilon_t}$ are the estimated first differenced residuals, $\hat{\varepsilon}_{t-1}$ are the estimated lagged residuals, ψ is the parameter of interest and u_t are the error terms. The hypothesis of the model can be restated as follow:

$H_0: \psi = 0$ (no stationarity)

$H_1: \psi < 0$ (stationarity)

At this point, it is possible to test the hypotheses on ψ as described in Section 2.7.2. The only difference with respect to the standard *ADF* test is that the critical values are different because the test is operating on residuals of an estimated model and not on an observed time series (Kirchgässner *et al.*, 2013, p. 217). The residuals have been constructed from a specific set of coefficient estimates with a specific sampling estimation error which change the distribution of the test statistic (Brooks, 2008, p. 339). The critical values differ depending on the number of variables tested for cointegration but also on the deterministic components of the equilibrium relationship (constant or linear time trend) and also on the number of observations. The most commonly used critical values for the *Augmented Dickey Fuller test* on the residual time series were computed by MacKinnon (1991) using Monte Carlo simulations. Comparing these critical values with those used for the standard *ADF* test, it is possible to notice that they are more negative meaning that more evidence against the null hypothesis is required in order to reject it and so it is less likely to allow the rejection of the null hypothesis. The null hypothesis of a unit root is rejected in favour of the alternative hypothesis if the test statistic is more negative than the critical value at some significance level, otherwise the null hypothesis cannot be rejected.

4. The final step concerns the estimation of the error correction model. It is possible to proceed to this stage only if the *ADF test* performed in the previous step indicates cointegration between the variables, i.e. $\widehat{\varepsilon_t} \sim I(0)$.

The error correction model can be described as (see Section 2.10.4):

$$\Delta x_t = \mu_1 + \alpha_1 z_{t-1} + \sum_{j=1}^{p} \gamma_{11}^{j} \Delta x_{t-j} + \sum_{j=1}^{p} \gamma_{12}^{j} \Delta y_{t-j} + \varepsilon_{1t}$$
$$\Delta y_t = \mu_2 + \alpha_2 z_{t-1} + \sum_{j=1}^{p} \gamma_{21}^{j} \Delta x_{t-j} + \sum_{j=1}^{p} \gamma_{22}^{j} \Delta y_{t-j} + \varepsilon_{2t}$$

where z_t is the process denoting the deviation from the long-term equilibrium. Since the actual deviation from the long-term equilibrium is unknown, Engle and Granger (1987) propose to substitute it with the residuals from the OLS regression performed in the second stage:

$$\begin{aligned} \Delta x_t &= \mu_1 + \alpha_1 \hat{\varepsilon}_{t-1} + \sum_{j=1}^{p} \gamma_{11}^{j} \Delta x_{t-j} + \sum_{j=1}^{p} \gamma_{12}^{j} \Delta y_{t-j} + \varepsilon_{1t} \\ \Delta y_t &= \mu_2 + \alpha_2 \hat{\varepsilon}_{t-1} + \sum_{j=1}^{p} \gamma_{21}^{j} \Delta x_{t-j} + \sum_{j=1}^{p} \gamma_{22}^{j} \Delta y_{t-j} + \varepsilon_{2t} \end{aligned} \qquad (2.50)$$

Then, the error correction equations (2.50) can be estimated separately using the OLS estimator. The shortfalls of the Engle-Granger method can be summarised as follow (Ssekuma, 2011):

- The estimation of the long-run equilibrium regression requires to choose the dependent variable and the independent variable. In the case of two variables, the Engle-Granger method can be applied by using the residuals from either of the following regression:

  $$y_t = \mu_1 + \beta_1 x_t + \varepsilon_{1t}$$

  or

  $$x_t = \mu_2 + \beta_2 y_t + \varepsilon_{2t}$$

  The theory suggests that a test for a unit root in the ε_{1t} sequence should be equivalent to the same unit root test in the ε_{2t} sequence if the sample size in sufficiently large. However, the properties of large samples may not be applicable to the actual sample sizes usually available. In practice, the problem is that it is possible to obtain ambiguous results with one regression indicating cointegration between the variables, and the other one indicating no cointegration.

- Engle- Granger methodology relies on a two-step estimator: first, the residual series ($\widehat{\varepsilon_t}$) is generated from the estimation of the long run equilibrium relationship between the time series using OLS. Then, the residual series is used to estimate the regression equation (2.49). Thus, the coefficient ψ is obtained by regressing the residuals from another regression on lagged differences of itself, meaning that any errors introduced in the first step are carried out also in the second step, making the results unreliable.

2.11.2 Johansen Methodology

An alternative method that could avoid the defects of the Engle-Granger methodology is the one introduced by Johansen (1988, 1991) which investigates cointegration in general multivariate systems where there are at least two integrated series. It is more powerful than the Engle-Granger methodology because it enables testing for the number of cointegrating vectors among *N* variables. However, it is important to highlight that the two methodologies have different objectives. The Engle-Granger method, being based on OLS, seeks the stationary linear combination that has the minimum variance, whereas the Johansen's method seeks the linear combination which is more stationary (Alexander, 2008, p. 235).

Johansen's procedure does not rely on OLS estimation, but it builds cointegrated variables directly on full-information maximum likelihood estimation, using sequential tests for determining the number of cointegrating vectors. Johansen tests can be thought of as a multivariate generalization of the *Augmented Dickey-Fuller* test described in Section 2.7.2 (Ssekuma, 2011).

The starting point of the Johansen' approach is the error correction model described in equation (2.46), from which the author formulates the following model of *N* variables:

$$\Delta y_t = \mu + \Gamma y_{t-1} + \sum_{j=1}^{p} \Gamma_j \Delta y_{t-j} + \varepsilon_t \qquad (2.51)$$

where $y_t = (y_{1t}, y_{2t}, \ldots, y_{Nt})'$ is a $N \times 1$ vector of N cointegrated variables; μ is a $N \times 1$ vector of constant terms; Γ_j are $N \times N$ matrices of coefficients; Γ is a $N \times N$ long-run coefficient matrix that can be interpreted as $\Gamma = \alpha\beta'$ where β is an $N \times h$ cointegrating matrix, with h cointegrating vector, and α is an $N \times h$ matrix of the adjustment coefficients, and finally $\varepsilon_t = (\varepsilon_{1t}, \varepsilon_{2t}, \ldots, \varepsilon_{Nt})$ is a $N \times 1$ vector of normally distributed disturbances.

The focus of Johansen's procedure is on the rank coefficient of y_{t-1}, which is the rank of the matrix Γ. The rank of a matrix is equal to the number of its characteristic roots, also called *eigenvalues*, that are different from zero.

Suppose that there are h cointegrating vectors, this means that the rank of the matrix Γ is h. The rank of the matrix, at least in principle, can range from 0 to *N-1*. Thus, the following cases should be considered (Kočenda and Černý, 2015, p. 171):

- If $h = 0$ then Γ is a zero matrix and all elements of y_t are integrated of order one, that is $y_t \sim I(1)$, and so equation (2.51) describes a Vector Auto-Regressive in first differences.
- If $0 < h \leq N - 1$ then all elements of y_t are integrated of order one, that is $y_t \sim I(1)$, h cointegrating vectors exist and equation (2.51) describes an error correction model.
- If $h = N$ then all elements of y_t are stationary, that is $y_t \sim I(0)$, and so equation (2.51) describes a Vector Auto-Regressive in levels.

The first phase of Johansen's approach concerns testing hypotheses about the rank of Γ. This test can be interpreted as a test for the number of its non-zero eigenvalues of the matrix Γ. Since Γ is an $N \times N$ matrix there will be N (theoretical) eigenvalues, denoted $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_N$. Suppose that there are $0 < h \leq N - 1$ cointegrating vectors, i.e. the rank of Γ is h, which means that there are *h* eigenvalues that are different from zero. In this context, it is possible to use the order sample of the estimated eigenvalues ($\hat{\lambda}_1 \geq \hat{\lambda}_2 \geq \ldots \geq \hat{\lambda}_N$) to test hypothesis about the rank of Γ. Johansen (1988, 1991) proposes two different likelihood ratio tests:

- The *trace test:* $\lambda_{trace}(h) = -T \sum_{i=h+1}^{N} ln(1 - \hat{\lambda}_i)$.
- The *maximum eigenvalue test:* $\lambda_{max}(h, h+1) = -T ln(1 - \hat{\lambda}_{h+1})$.

Here *T* is the sample size and $\hat{\lambda}_i$ is the i^{th} estimated eigenvalue. The *trace test* is used to test the null hypothesis that the rank of Γ is less than or equal to h, i.e. $H_0: rank\ \Gamma \leq h$, against the alternative hypothesis that the rank of Γ is greater than *h*, i.e. $H_1: rank\ \Gamma > h$. The *maximum eigenvalue test* is used to test the null hypothesis that the rank of Γ is less than or equal to *h*, i.e. $H_0: rank\ \Gamma \leq h$, against the alternative hypothesis that the rank of Γ is *h+1*, i.e. $H_1: rank\ \Gamma = h + 1$. These tests are likelihood ratio (see Hamilton, 1994, pp. 144-145) but they do not have the usual chi-squared distributions. As in the unit root tests, *Augmented Dickey-Fuller test* (see Section 2.7.2) and *Phillips-Perron test* (see Section 2.7.3), the percentiles of the distributions depend on whether a constant or a time trend are included in the model specification. Moreover, in this case, the percentile distributions depend on the value of $N - h$, that is the number of nonstationary components under the null hypothesis, which is given by the number of variables described by the Vector Autoregressive model (*N*) minus the number of cointegration relations under the null hypothesis (h). Critical values for the *trace test* and for the *maximum eigenvalue test* can be tabulated using *Monte Carlo simulation*, as in Osterwald-Lenum (1992). A set of critical values for both tests is reported in *Appendix A* (*Table A4* and *Table A5*).

Consider the *trace test*, if the test statistic is greater than the corresponding critical value for a given level of significance then the null hypothesis that there are h cointegrating vectors (or less) is rejected in favour of the alternative hypothesis that there are more than h cointegrating vectors. Similarly, if the value of the *maximum eigenvalue test* is greater than the corresponding critical value for a given level of significance then the null hypothesis that there are h cointegrating vectors (or less) is rejected in favour of the alternative hypothesis that there are $h+1$ cointegrating vectors. Thus, conducting an appropriate sequence of tests with $\lambda_{max}(h, h+1)$ or $\lambda_{trace}(h)$ and considering different values of h, it is possible to estimate the rank of the matrix Γ. At this point, given the rank of the matrix and imposing the reduced rank restrictions in $\Gamma = \alpha\beta'$, maximum likelihood estimates of model parameters can be calculated (see Hamilton, 1994, pp.637-638). It is important to notice that the parameters α and β are not uniquely identified, since different combinations of these parameters produce the same matrix $\Gamma = \alpha\beta'$. In other words, using full-information maximum likelihood estimation (see Hamilton, 1994, pp. 247-250) it is only possible to estimate the linear space spanned by the cointegrating vectors. Thus, the cointegrating vectors contained in β have to be normalized in order to obtain unique cointegrating relations. The choice concerning the identification of an adequate normalization of the cointegrating vector cannot be based on a rule of thumb, but it has to be made on the basis of an economic interpretation of the data considered.

Chapter 3

Pairs Trading using Cointegration

3.1 Pairs Trading

The strategy of statistical pairs trading was first pioneered in the mid 1980's by a quantitative trading group headed by Nunzio Tartaglia at Morgan Stanley. They found that certain pairs of securities were correlated in their day-to-day price movements, i.e. the prices of certain pairs of securities tended to move together over time. However, they noticed that sometimes these relationships presented some anomalies, and whenever such an event occurred the pair was traded based on the assumption that the anomaly would correct itself in the future leading to a profit (Vidyamurthy, 2004, pp. 73-74). The key idea behind the concept of pairs trading is that it is possible to gain a profit by exploiting temporary anomalies between prices of related securities which have a long-term equilibrium. When such an event occurs, one security will be overvalued relative to the other one, or conversely one security will be undervalued relative to the other. This mutual mispricing between the two securities' prices is captured by the notion of spread. Consider two stocks A and B and their prices at time *t*, which are respectively p_t^A and p_t^B, their spread at time *t* can be defined in two different ways:

1. As the scaled difference in the price of the two stocks, that is:

$$Spread_t = p_t^A - \gamma p_t^B \tag{3.1}$$

2. As the scaled difference of the logarithm of the two stocks' prices, that is:

$$Spread_t = log(p_t^A) - \gamma log(p_t^B) \tag{3.2}$$

where γ is a constant representing the scaling factor.

In the pairs trading framework, the spread can be thought as the degree of mutual mispricing between two related stocks: the greater the spread, the higher the magnitude of mispricing and the greater the profit potential.

Pairs trading involves the creation of a pair portfolio, where the overvalued security is sold (short position) and the undervalued security is bought (long position), whenever the spread is away from its long-term equilibrium value. A profit can be made by closing out the trade, i.e. by selling the long position and off-setting the short position, upon convergence of the spread to its equilibrium value. In other words, pairs trading is designed to capture profits by exploiting short-term deviations from a long-run equilibrium between securities.

Whenever a pairs trading strategy is constructed and implemented properly, the underlying behaviour of the stock market does not affect the profits of the pair portfolio, i.e. the profits generated by the portfolio should be independent of the general stock market returns (Dunis and Ho, 2005). Thus, as discussed in Section 1.1, pairs trading can be considered a market-neutral trading strategy since the trader profits simply by exploiting short-term discrepancies in the prices of the securities regardless of whether the market goes up or down. Market-neutral trading strategies combine a long position with a short position in order to reduce directional exposure, and so they are hedged against the market risk (Miao, 2014).

Nevertheless, it is must be pointed out that pairs trading classifies as statistical arbitrage strategy rather than a pure arbitrage one (see Section 1.1), and so it is not riskless. Recalling the definition proposed by Lazzarino *et al.* (2018), statistical arbitrage qualifies as a relative value strategy with positive expected excess return and a tolerably small potential loss (see Section 1.1). From this definition it follows that the risk of potential losses, due for example to market events, structural price changes or persistent pricing inefficiencies (Lin *et al.*, 2006), is an important component of pairs trading strategies. By using statistical tools, this risk needs to be assessed and compared with the expected positive excess return in order to evaluate the strategy's profitability.

According to Chen *et al.* (2019) pairs trading follows a two-step process. First, identification of pairs of trading instruments whose prices are found to be highly correlated, i.e. the price of one instrument moves in the same direction as the other, during a formation period. Second, monitor the spread between them in a subsequent trading period, and when a sufficient divergence in the spread is observed, a long-short position is simultaneously established.

Pairs trading relies on the key assumption of the existence of a long-term equilibrium in the spread. Put differently, pairs trading requires the spread to be a stationary process (see section 2.2) so that any deviation from the equilibrium has only a transitory effect and tends to disappear over time, bringing the spread back to its long-term value (see Section 2.10.1).

The main shortfall of the aforementioned traditional two-step process is that the identification process is based on correlation rather than on cointegration. According to Alexander and Dimitriu (2005) correlation analysis is only valid for stationary variables. Since most macroeconomic variables, such as prices, are usually found to be nonstationary, this type of analysis in order to be implemented requires prior de-trending of the variables. The main weakness resulting from this procedure is that many valuable information is lost. In particular, any long-term trend is removed from the data so that it is not possible to base any decision on common trends in prices. By contrast, the main purpose of cointegration analysis is to test whether the prices share any common stochastic trends; if it does exits then the prices must have a long term equilibrium relationship (see Section 2.10.3), which can be used in an Error

Correction Model (see Section 2.10.4) to explain how short-term deviations from the equilibrium value are corrected (Alexander, 2008, pp. 225-226).

According to Alexander (1999), cointegration and correlation are related but very different concept. Correlation is intrinsically a short run measure since it reflects co-movement in returns, which are usually unstable over time. For this reason, correlation based portfolio allocation strategies or hedging strategies require frequent rebalancing. Furthermore, long-short strategies, such as pairs trading, that are based only on correlation cannot promise long term performance because there is not a mechanism guaranteeing the existence of a long-term equilibrium value and the subsequent reversion to it. On the other hand, cointegration measures long-term co-movements in prices. Therefore, cointegration enables the trader to combine the stocks in a certain linear combination so that the resulting portfolio is stationary. If the combination of the two stocks is stationary, this means that they share a long-term equilibrium relationship to which they will tend to converge over time (see Section 2.10.4). In other words, cointegration incorporates mean reversion into pairs trading framework, which is the most important statistical relationship necessary for a profitable strategy: if the value of a portfolio is known to fluctuate around an equilibrium value, any deviations from it can be exploited to gain a profit (Puspaningrum, 2012).

The objective of this chapter is to expand the traditional two-step process based on correlation in order to provide a structure for the design and analysis of pairs trading using cointegration.

Vidyamurthy (2004, pp. 83-84) provides the most cited work for cointegration-based pairs trading. The framework relies on three crucial steps.

1. The first step, examined in Section 3.2, concerns the identification of stock pairs that could potentially be cointegrated.
2. The second step, discussed in Section 3.3, concerns the verification of the hypothesis that the stock pairs identified in the previous point are indeed cointegrated, based on the statistical analysis of the data.
3. The third step, considered in Section 3.4, concerns the design of proper trading rules, which can be based on parametric or nonparametric approaches.

Finally, a literature review of the cointegration-based pairs trading strategy is examined in Section 3.5.

3.2 Pre-Selection of Stock Pairs

The first phase of cointegration based pairs trading consists in the identification of potential stock pairs. The aim of this procedure is to short-list the candidates of pairs trading in order to reduce the pairs for cointegration testing and further analysis. Pre-selection techniques must be simple and straightforward relative to cointegration testing, otherwise it would be more convenient to test directly for cointegration with all the pairs from the universe of stocks considered.

The approach suggested by Vidyamurthy (2004, p.86) aims at producing an ordered list of the pairs based on the degree of co-movement. For each pair is computed a score (distance measure): the higher this value, the greater the degree of co-movement.

The methodology proposed by Vidyamurthy (2004) exploits the common trend model, introduced by Stock and Watson (1988), to motivate the use of a distance measure based on correlation as a pre-selection technique to rank the candidates of pairs trading. As discussed in Section 2.10.3, the idea behind the common trend model is that a time-series can be expressed as the sum of a stationary component and a nonstationary component:

$$x_t = \eta_{x_t} + \varepsilon_{x_t}$$

$$y_t = \eta_{y_t} + \varepsilon_{y_t}$$

where $\{\eta_{x_t}\}$ and $\{\eta_{y_t}\}$ represent the random walk components (common trends component), while $\{\varepsilon_{x_t}\}$ and $\{\varepsilon_{y_t}\}$ represents the stationary components (specific components). The two time series are cointegrated if and only if there exists a parameter γ, which can assume any real value, so that their stochastic trends are generated by the same random walk process and can only differ by a linear scaling factor γ, which represents the cointegration coefficient, such that:

$$\eta_{x_t} = \gamma\eta_{y_t} \tag{3.3}$$

Consider the innovation sequences ($r_{x_t}$and r_{y_t}) resulting from the random walk component of the time series, which can be obtained simply by differencing them:

$$r_{x_{t+1}} = \eta_{x_{t+1}} - \eta_{x_t}$$

$$r_{y_{t+1}} = \eta_{y_{t+1}} - \eta_{y_t}$$

According to equation (3.3), the random walk components must be identical up to a scalar, so it must also be true that the innovation derived from those random walk components must be identical up to the same scalar, so that:

$$r_{x_{t+1}} = \gamma r_{y_{t+1}} \tag{3.4}$$

According to Vidyamurthy (2004, p. 88), if two variables are identical up to a scalar, they must be perfectly correlated, which means that if the cointegration coefficient is positive (negative), then the correlation coefficient must be +1 (-1). Put differently, if two time series are cointegrated, the innovation sequences derived from the random walk components must be perfectly correlated.

Furthermore, given a linear relationship between innovation sequences, such as the one described in equation (3.4), the cointegration coefficient can be estimated by performing a simple regression of one innovation sequence against the other (Vidyamurthy, 2004, p. 89):

$$\hat{\gamma} = \frac{cov(r_{x_{t+1}}, r_{y_{t+1}})}{var(r_{y_{t+1}})} \quad (3.5)$$

According to Vidyamurthy (2004, pp. 90-92), the rationale of using correlation as a selection filter comes from arbitrage pricing theory (APT), which suggests that stock returns may be decomposed into common factor returns, which represent the returns based on the exposure of stocks to diverse risk factors, and idiosyncratic returns, which represent the specific components of the stocks . APT states that two assets with the same risk factor exposure profile should have the same expected return (see Vidyamurthy, 2004, pp. 39-42).

The idea proposed by Vidyamurthy (2004, pp. 90-92) is that the common factor returns of the APT theory can be interpreted as the innovations derived from the common trends model. To clarify this concept, consider two stocks A and B with the following common risk factor returns vectors:

Stock A $\quad w_t = (w_{1,t}, w_{2,t}, \dots, w_{n,t})$

Stock B $\quad \gamma w_t = (\gamma w_{1,t}, \gamma w_{2,t}, \dots, \gamma w_{n,t})$

which are identical up to a scalar (γ). Let $b = (b_1, b_2, \dots, b_n)$ be the vector of factor loadings, i.e. the vector describing the exposure of a stock return with respect to the n common risk factors, and $\{r_{A,t}^{spec}\}$ and $\{r_{B,t}^{spec}\}$ be the idiosyncratic returns for the two stocks which must be stationary. The returns for the stocks can be expressed as:

$$r_{A,t} = (b_1 w_{1,t}, b_2 w_{2,t}, \dots, b_n w_{n,t}) + r_{A,t}^{spec}$$

$$r_{B,t} = \gamma(b_1 w_{1,t}, b_2 w_{2,t}, \dots, b_n w_{n,t}) + r_{B,t}^{spec}$$

where

$$r_{A,t}^{cf} = (b_1 w_{1,t}, b_2 w_{2,t}, \dots, b_n w_{n,t})$$

$$r_{B,t}^{cf} = \gamma(b_1 w_{1,t}, b_2 w_{2,t}, \dots, b_n w_{n,t})$$

are the common factor returns for the two stocks. Since the innovation sequences of the common trend are identical up to a scalar, i.e. $r_{B,t}^{cf} = \gamma r_{A,t}^{cf}$, the condition for cointegration is satisfied.

Finally, consider the following linear combination of the returns:

$$r_{B,t} - \gamma r_{A,t} = (r_{B,t}^{cf} - \gamma r_{A,t}^{cf}) + (r_{B,t}^{spec} - \gamma r_{A,t}^{spec})$$

Assuming that the stocks are cointegrated, i.e. $r_{B,t}^{cf} = \gamma r_{A,t}^{cf}$, then the return of the portfolio composed by a long position in one share of stock B and a short position in γ shares of A becomes:

$$r_{B,t} - \gamma r_{A,t} = r_{B,t}^{spec} - \gamma r_{A,t}^{spec}$$

Which means that the return of the portfolio depends only on specific returns.

Since the common factor returns of the Arbitrage Pricing Theory can be interpreted as the innovations derived from the common trends model, the necessary condition for cointegration, discussed above, can be restated as follows: if the common factor returns of the stocks are perfectly correlated thein it is possible to conclude that they are cointegrated. The distance measure proposed by Vidyamurthy (2004, p. 94) is the absolute value of the correlation of the common factor return:

$$|\rho| = \left| \frac{cov(r_A^{cf}, r_B^{cf})}{\sqrt{var(r_A^{cf})\, var(r_B^{cf})}} \right|$$

The closer this value is to one, the greater the degree of co-movement.

Based on this value, the trader can create a ranking of all the potential pairs and short-list them so that only the pairs with a distance measure sufficiently close to one will be tested for cointegration.

The model proposed by Vidyamurthy (2004) suggests that the cointegration coefficient (γ) should be interpreted as the relative risk factor exposure in the two stocks, so that one share of stock B exposes the trader to the same amount of systemic risk as γ shares of stock A. Do *et al.* (2006) criticize this argument by pointing out that under the APT, the return due to exposure to risk factor is on top of the risk free return ($r_{f,t}$):

$$r_{A,t} = r_{f,t} + (b_1 w_{1,t}, b_2 w_{2,t}, \dots, b_n w_{n,t}) + r_{A,t}^{spec}$$

$$r_{B,t} = r_{f,t} + \gamma(b_1 w_{1,t}, b_2 w_{2,t}, \dots, b_n w_{n,t}) + r_{B,t}^{spec}$$

Thus, it is not generally true that when the risk exposure profiles of two stocks are identical up to a scalar the return of one share of stock B is identical to the return of γ shares of stock A plus some Gaussian noise, as suggested by Vidyamurthy (2004). However, in order to overcome the problem pointed out by Do *et al.* (2006) it is sufficient to consider in the procedure introduced by Vidyamurthy (2004) the excess returns, defined as $(r_{A,t} - r_{f,t})$ and $(r_{B,t} - r_{f,t})$, instead of the returns, defined as $r_{A,t}$ and $r_{B,t}$.

Miao (2014) proposes two different methodologies for ranking and selecting stock pairs: the first is based, similarly to the distance measure, on the correlation coefficients, while the second is based on the minimum-distance criterion.

Consider two stocks A and B and their prices at time *t* which are respectively p_t^A and p_t^B. The first approach proposed by Miao (2014) is based on the Pearson correlation coefficient (ρ) of the two stocks, which can be computed as follow:

$$\rho = \frac{\sum_{t=1}^{T}(p_t^A - \overline{P^A})(p_t^B - \overline{P^B})}{[\sum_{t=1}^{T}(p_t^A - \overline{P^A})^2 \sum_{t=1}^{T}(p_t^B - \overline{P^B})^2]^{1/2}}$$

where T represents the sample size considered, while $\overline{P^A}$ and $\overline{P^B}$ are respectively the sample mean of the prices of stocks A and B.

According to Miao (2014) the closer ρ is to one, the more stocks are correlated, which means that stocks A and B are highly matched pairs. Thus, only pairs with a sufficiently high value of ρ will be selected for the cointegration testing phase. It is important to highlight that this methodology is based on prices, and so it is substantially different from the distance measure based on correlation proposed by Vidyamurthy (2004), which is based on returns.

However, according to Chan (2013, p.65) since it is normally the case that log-prices are cointegrated when the prices are cointegrated, both the choice to use prices and log-prices are theoretically justified.

The minimum-distance criterion (SSD) (see Section 1.2) consists in the sum of squared deviations between normalized stock prices:

$$SSD_{A,B} = \sum_{t=1}^{T}(P_t^A - P_t^B)^2$$

where P_t^A and P_t^B are respectively the normalized price of stocks A and B at time *t*.

The smaller the deviation between the normalized prices, i.e. the closer the value of $SSD_{A,B}$ is to 0, the more similar the two stocks will be. Thus, only pairs with a sufficiently small value of $SSD_{A,B}$ will be chosen for the next stage.

As highlight by Miao (2014) the correlation coefficient and the minimum-distance criterion can be used only as pre-selection criteria. Indeed, as discussed in Section 3.1, a high level of correlation is not sufficient to ensure mean reversion between prices, since correlations is a short-run measure which is very unstable over time.

So far, quantitative methods for selecting stock pairs have been analysed. However, qualitative methods, based on economic reasons, for short-listing candidates of pairs trading can also be considered, such as stocks belonging to the same industry. The intuition is that it is more probable to find cointegrated price stock series within the same sector because they are exposed to the same market risks and are affected by the same driving factors. The main disadvantage of this method is that it does not provide any criteria to rank all the potential stock pairs. As a consequence, if the sector considered is very big, this approach will prove to be of little use in the identification of potentially cointegrated pairs. In this case, a solution could be to combine this methodology with one of the others described above.

3.3 Testing for Tradability

The previous section discusses the procedures that can be implemented to select potential stocks for pairs trading. The objective of this section is to verify if the identified stocks are actually cointegrated.

Vidyamurthy (2004, p. 105), in order to test for cointegration, adopts the Engle-Granger approach (see Section 2.11.1), which is a simple residual-based testing methodology based on two main steps:

1. Ordinary Least Squares estimation of the long run equilibrium relationship between the time series considered.
2. Application of unit root tests on the estimated residuals of the OLS regression, in order to verify whether they are stationary or not.

Vidyamurthy (2004, p. 106) considers the case in which the *log* price of stock A, $log(p_t^A)$, is regressed against the *log* price of stock B, $log(p_t^B)$:

$$log(p_t^A) - \gamma log(p_t^B) = \mu + \varepsilon_t \quad (3.6)$$

where γ is the cointegration coefficient, μ represents the equilibrium value, which captures some sense of “premium” in stock A versus stock B (Do *et al.*, 2006), and ε_t is a time series with mean 0 and variance σ_ε^2, which represents the disturbance term in the equilibrium.

Recalling equation (3.5), the cointegration coefficient in equation (3.6) can be estimated by performing a regression of the common factor returns of one stock, against the other. Notice that there are two possible values of the cointegration coefficient depending on the choice of the independent variable (see Section 2.11.1):

1. The linear relationship is expressed considering stock B to be the independent variable and stock A to be the dependent variable:

$$\widehat{\varepsilon_t} = log(p_t^A) - \hat{\gamma} log(p_t^B)$$

$$\hat{\gamma} = \frac{cov(r_{A,t}^{cf}, r_{B,t}^{cf})}{var(r_{B,t}^{cf})}$$

where $r_{A,t}^{cf}$ and $r_{B,t}^{cf}$ are the common factor returns, defined as log-prices, at time *t* of stock A and stock B, respectively.

2. The linear relationship is expressed considering stock A to be the independent variable and stock B to be the dependent variable:

$$\widehat{\varepsilon_t} = log(p_t^B) - \widehat{\gamma'} log(p_t^A)$$

$$\widehat{\gamma'} = \frac{cov(r_{A,t}^{cf}, r_{B,t}^{cf})}{var(r_{A,t}^{cf})}$$

In order to choose between the two alternative estimates of the cointegration coefficient, Vidyamurthy (2004, p. 108) suggests choosing the larger one because it is the one with the lower variance. In fact, if $\widehat{\gamma'} > \hat{\gamma}$ then it follows that $var(r_{A,t}^{cf}) < var(r_{B,t}^{cf})$.

Finally, the estimated residual series ($\widehat{\varepsilon_t}$) is tested for stationarity using the Augmented Dickey Fuller test (see Section 2.7.2). If the residuals are found to be stationary, i.e. if the ADF test rejects the null hypothesis of no stationarity, then $log(p_t^A)$ and $log(p_t^B)$ are cointegrated.

At this point, it is possible to rank all the stock pairs based on two different criteria:

1. The *cointegration test values*: this approach is based on the value of the ADF test (or any other unit root test) used to verify the stationarity of the residuals: the smaller the value of the test, the higher the rank of the pair (Miao, 2014).
2. The *Sharpe ratio* (SR): it measures the risk-adjusted returns of a pair of stocks and is calculated as:

$$SR = \frac{\overline{r_p} - r_f}{\sigma_p}$$

where $\overline{r_p}$ is the mean return associated to the pair, r_f is the risk-free rate and σ_p is the standard deviation associated to the pair. In this case, the higher the Sharpe ratio, the higher the rank of the pair (Caldeira and Moura, 2013).

Equation (3.6) can be interpreted as the return on a portfolio consisting of a long position in one share of stock A and a short position in γ shares of stock B. Cointegration between the two stocks implies that the spread time series, i.e. $log(p_t^A) - \gamma log(p_t^B)$, has a long run mean μ and any deviations from this value are only temporary fluctuations, since $\varepsilon_t \sim I(0)$ (Do *et al.*, 2006).

Do *et al.* (2006) criticised the approach proposed by Vidyamurthy (2004) mainly for two reasons: first, the Engle-Granger methodology makes results sensitive to the ordering of the variables, and so it creates a potential problem of ambiguity where one regression may indicate cointegration and the other one may indicate no cointegration (see Section 2.11.1). Second, if the bivariate series considered are not cointegrated, the cointegrating regression, described in equation (3.6), could lead to spurious regression (see Section 2.9).

Many researchers, such as Huck and Afawubo (2015) and Caldeira and Moura (2013), in order to avoid the asymmetry problem in treating variables, perform a more rigorous test of cointegration, which is the Johansen test based on a Vector Error Correction Model (VECM). As discussed in Section 2.11.2, the Johansen test is more informative than the Engle-Granger because it investigates cointegration in general multivariate systems where there are at least two integrated series, allowing to test for the number of cointegrating vectors at the same time.

However, as long as only pairs of potentially cointegrated stocks are considered, there exist only one cointegrating vector which can be uniquely identified normalizing to one the coefficient assigned to one of the two stocks (see Section 2.10.2). Thus, in this type of framework, the Johansen test is not necessarily more powerful than the Engle-Granger methodology.

Furthermore, according to Alexander (2008, p. 239) there could be at least two good reasons for choosing the Engle-Granger as the preferred methodology for some financial application. First, from a risk management prospective, the criterion of minimum variance (typical of the Engle-Granger approach) is often more important than the criterion of maximum stationarity (typical of the Johansen approach). Second, there is often a natural choice of dependent variables in the cointegrating regressions, which eliminates the ambiguity problem discussed above.

3.4 Trading Design

The final stage of the framework developed by Vidyamurthy (2004) for cointegration-based pairs trading concerns the design of optimal trading rules, i.e. optimal entry/exit thresholds, which are based on the profit maximization principle.

The model used is the one described in equation (3.6), which is reported below:

$$log(p_t^A) - \gamma log(p_t^B) = \mu + \varepsilon_t$$

The basic trading idea is to open a spread position on a deviation of Δ from the equilibrium value and revert the position upon mean reversion. In other words, the strategy is to open a long position in the spread portfolio, i.e. buy one share of stock A and sell γ shares of stock B, when it is sufficiently below its long-run equilibrium ($\mu - \Delta$), and to short the spread portfolio, i.e.

sell one share of stock A and buy γ shares of stock B, when it is sufficiently above its long-run equilibrium ($\mu + \Delta$). Once the portfolio mean reverts to its equilibrium value then position is unwound and a profit is obtained (Puspaningrum, 2012).

The main objective of the trading design phase is to find the value of Δ that maximises the profit function, which can be expressed as the product between the profit per trade and the numbers of trade. According to Vidyamurthy (2004, p. 124), any choice for the threshold level has a profit per trade associated with it. If one is able to calculate the rate at which the threshold level is crossed (rate of zero crossing), it is possible to determine the expected number of trades. Thus, the total expected profit can be calculated simply by multiplying the profit per trade with the expected number of trades. This approach can be repeated for different threshold levels, and the value of Δ which yields the higher profit is chosen as optimal trigger.

The main problem of this methodology is that estimating the rate of zero crossing is not an easy task. Vidyamurthy (2004, p. 125) considers the case of a spread modelled as an ARMA process, for which the rate of zero crossing can be calculated using the Rice's formula (Rice, 1944).

An alternative approach, proposed by Puspaningrum (2012), suggests using the first passage-time (see Elliott *et al.* (2005), Bertram (2010)) for stationary series which calculates the time needed for the time series to mean revert to its long-run equilibrium after crossing a pre-specified threshold level (see Section 1.4).

In practice, the most used technique to determine when to open and when to close a position is based on a standard deviation metric, as in the distance method proposed by Gatev *et al.* (2006) (see Section 1.2). For each cointegrated pairs identified during the previous steps (see Section 3.2 and Section 3.3), the spread at time *t* is defined as the scaled difference of the logarithm of the two stocks' prices:

$$Spread_t = log(p_t^A) - \gamma log(p_t^B)$$

Since the two stocks are cointegrated, the spread is a stationary time series, that is:

$$Spread_t \sim I(0)$$

At this point, the dimensionless *z*-score (or normalized spread), which measures the distance to the long-run mean in units of long-term standard deviation (Caldeira and Moura, 2013), is calculated as:

$$z_t = \frac{Spread_t - \mu_e}{\sigma_e}$$

where μ_e is the spread's mean and σ_e is the spread's standard deviation, both calculated using the data of the formation period considered.

Most authors, such as Rad *et al*. (2016) and Caldeira and Moura (2013), consider the 2-standard deviation rule (2 and 3-standard deviation rule is considered by Huck and Afawubo (2015) (see Section 3.5.2)), introduced by Gatev *et al.* (2006), as opening trigger. According to this rule, it is possible to identify two different opening thresholds:

- *Lower threshold*: when the *z*-score is less than or equal to -2, this means that the portfolio of pairs is sufficiently below its long-run equilibrium. In this case, the portfolio is undervalued and so one should purchase it, which means simultaneously buy one share of stock A and sell γ shares of stock B.
- *Upper threshold*: when the *z*-score is greater than or equal to 2, this means that the portfolio of pairs is sufficiently above its long-run equilibrium. Thus, since the portfolio is overvalued one should sell it, which means simultaneously sell one share of stock A and buy γ shares of stock B.

It is important to highlight that the greater the trigger, i.e. the greater the deviation from the long-run equilibrium required to open a trade (e.g. 3-standard deviation rule), the lower the number of openings and trades during a specific trading period. However, a higher opening threshold would yield a higher profit per trade than a lower value. On the other hand, a lower threshold-value will lead to more trades during a specific trading period, potentially increasing the total profits. Thus, it is not possible to determine a priori whether total profits increase or decrease with higher thresholds value.

Finally, the position is closed when the *z*-score approaches zero again, which translates into the pair returning to their long-term equilibrium. However, it is not necessary that a position is closed when $z_t = 0$; for example, Caldeira and Moura (2013) select empirically slightly different exit signals, following a procedure already used by Avellaneda and Lee (2010), specifically:

- A short position in the portfolio is closed when $z_t < 0.75$.
- A long position in the portfolio is closed when $z_t > -0.50$.

According to Lin *et al.* (2006), the main risks of the aforementioned trading strategy are that, due for example to significant pricing inefficiencies or particular market events:

- Price spreads may continue to diverge after position opening rather than revert to the long-term equilibrium, so that when the trader is forced to close the trade on the last day of the trading period, he/she might experience substantial losses.
- The long-run equilibrium value may vary during the trading period, so that the developed trading strategy becomes completely unreliable, leading to potential losses for the trader.

The profit reduction consequences of these risks can be offset with the introduction of some additional rules, aimed at limiting the loss of too much trading capital on a single pairs trading, such as:

- A *stop-loss constrain*, which is a function used to automatically unwind a position whenever a pre-defined loss is registered (e.g. Caldeira and Moura (2013) set a stop loss to close a position if a loss of 7% is observed).
- *Maximum holding length of a trade*: which is the maximum time a trade can be kept opened, exceeding this generates an exit signal (e.g. Caldeira and Moura (2013) set the maximum holding length of a trade to 50 days, because according to their data the average profitability of the strategy starts decreasing after this period of time).

Summing up

The standard procedure for the implementation of a pairs trading strategy based on cointegration can be summarized as follows (Caldeira and Moura, 2013):

1. The data considered are initially divided into formation (usually one or two years) and trading periods (usually 4 or 6 months). During the formation period (or training period) the parameters of the experiment are computed. During the trading period (or testing period) the experiment is run using the parameters computed in the training period.
2. During the formation period, all the possible combinations of pairs are short-listed based on the pre-selection procedures discussed in Section 3.2. The pairs identified are tested for cointegration using the Engle-Granger approach or the Johansen test. Finally, the pairs that passed the cointegration test are ranked based on the cointegration test values or based on their Sharpe ratio (see Section 3.3).
3. The best pairs identified during the formation period are used, during the trading period, to test the performance of the pairs trading strategy.
4. At the end of each trading period the positions that were opened are closed, and a new training period starting on the last observation of the previous trading period is initiated. This procedure continues in a rolling window fashion until the end of the sample considered.

According to Puspaningrum (2012), there is not a standard rule for deciding the lengths of the formation period and the trading period. The formation period has to be long enough so that it is possible to verify whether a cointegration relationship exist or not, but not so long that there is not enough information for the following trading period. The trading period is chosen so that the selection process is recent, and round-trips have time to occurs using reasonable opening triggers, but not so long because it is possible that the cointegration relationship between two stocks may change over time.

3.5 Literature Review

This section presents three interesting applications of the cointegration-based pairs trading strategy which can be found in the literature. In Section 3.5.1, we examine the study by Lin *et al.* (2006) which develop a minimum profit condition for a pair of cointegrated securities. In Section 3.5.2, we analyse the work by Huck and Afawubo (2015) which run a comparison study, analysing the cointegration approach, the distance approach, and the stationarity of the price ratio approach (this approach is based on the idea that in order to generate profits in a pair-trade, the price ratio between two stocks has to be a stationary process) for the S&P 500 constituents and using different parametrizations, in order to understand if one of these selection methods dominates the other in terms of monthly returns. Finally, in Section 3.5.3, we consider the research by Naccarato *et al.* (2019) which aims at solving the Markowitz portfolio optimization problem through the pairs trading cointegrated strategy.

3.5.1 Loss Protection in Pairs Trading Through Minimum Profit Bounds: A Cointegration Approach (Lin *et al.*, 2006)

The strategy proposed by Lin *et al.* (2006) is based on the so called *Cointegration Coefficient Weighted* (CCW) *Rule*. The idea behind the CCW rule is to trade a pair of stocks which are found to be cointegrated, based on the cointegration coefficient, in order to achieve a guaranteed minimum profit per trade.

The strategy developed by Lin *et al.* (2006) is based on the following assumptions:

- The two share price series are cointegrated over the entire time horizon considered.
- Stock *A* always represents the short position (sell), while stock *B* always represents the long position (buy).
- At the opening of any trade, based on the relationship described in equation (3.7), the price received for one share of stock A (short position) is always higher than the price paid for γ shares of stock B (long position), i.e. $p_{A,t} > \gamma p_{B,t}$.
- It is possible to open a trade only if the previous opened trade is closed yet.

Consider two stocks *A* and *B* whose prices are integrated process of order 1, i.e. $p_{A,t} \sim I(1)$ and $p_{B,t} \sim I(1)$. According to the first assumption the two share price series are cointegrated over the period considered, so there must be a non-zero linear combination that is stationary, so that:

$$p_{A,t} - \gamma p_{B,t} = \mu + \varepsilon_t \tag{3.7}$$

where $\{\varepsilon_t\}$ is the cointegration errors and it is a stationary time series, i.e. $\varepsilon_t \sim I(0)$, μ represents the equilibrium value, and $\gamma > 0$ is the cointegration coefficient. Notice that the cointegrating relationship described in (3.7) is different from the one described in (3.6) because: in (3.7) the spread is described as a scaled difference in the price of the two stocks, while in (3.6) the spread is described as a scaled difference of the logarithm of the two stocks' prices.

In order to ensure that the proceeds from the sale of stock A at time t_0 (time at which a trade is opened) are sufficient to cover the outflow to buy stock B, the following condition needs to be satisfied:

$$N_A p_{A,t_0} \geq N_B p_{B,t_0} \tag{3.8}$$

where N_A and N_B are the number of shares in the short position and in the long position at time t_0, respectively.

Open trade condition (OTC(φ))

A time t_0 can be considered as an open trading time if it satisfies the following condition:

$$p_{A,t_0} - \gamma p_{B,t_0} = \mu + \varepsilon_{t_0} > \mu + \varphi \tag{3.9}$$

where φ is a positive real number, representing the opening trigger.

At this point, one can calculate the total profit (π_{t_C}) per trade obtained at time t_C (time at which the trade is closed) as follow:

$$\pi_{t_C} = N_B[p_{B,t_C} - p_{B,t_0}] + N_A[p_{A,t_0} - p_{A,t_C}] \tag{3.10}$$

To ensure that both condition (3.8) and (3.9) are satisfied, another condition on N_A and N_B needs to be established. If a trader decides to buy N_B shares of stock *B*, then he/she must sell at least ${N_B}/{\gamma}$ shares of stock *A*, that is $N_A = {N_B}/{\gamma}$. Substituting this condition in equation (3.10), one gets:

$$\pi_{t_C} = N_B[p_{B,t_C} - p_{B,t_0}] + \frac{N_B}{\gamma}[p_{A,t_0} - p_{A,t_C}] \tag{3.11}$$

The first term in equation (3.11) can be rewritten, using equation (3.9), as follow:

$$N_B[p_{B,t_C} - p_{B,t_0}] = N_B\left[\frac{1}{\gamma}(p_{A,t_C} - \mu - \varepsilon_{t_C}) - \frac{1}{\gamma}(p_{A,t_0} - \mu - \varepsilon_{t_0})\right]$$
$$= \frac{N_B}{\gamma}\left[p_{A,t_C} - p_{A,t_0} + \varepsilon_{t_0} - \varepsilon_{t_C}\right] \quad (3.12)$$

Substituting equation (3.12) in equation (3.11), one obtains:

$$\pi_{t_C} = \frac{N_B}{\gamma}\left[p_{A,t_C} - p_{A,t_0} + \varepsilon_{t_0} - \varepsilon_{t_C}\right] + \frac{N_B}{\gamma}\left[p_{A,t_0} - p_{A,t_C}\right]$$
$$= \frac{N_B}{\gamma}\left[p_{A,t_0} - p_{A,t_C} - \left(\varepsilon_{t_C} - p_{A,t_C}\right) + \left(\varepsilon_{t_0} - p_{A,t_0}\right)\right] \quad (3.13)$$
$$= \frac{N_B}{\gamma}(\varepsilon_{t_0} - \varepsilon_{t_C})$$

<u>*Close trade condition (CTC (φ, ω))*</u>

The final step concerns the identification of an appropriate closing time t_C, such that a trader who opens a trade under *OTC(φ)*, by buying N_B shares of stock *B* and selling ${N_B}/{\gamma}$ shares of stock *A*, will be able to gain a minimum profit of *K* when the trade is closed. From equation (3.13), to ensure that a minimum profit per trade (*MPPT*) of *K > 0* is gained, the following condition needs to be satisfied:

$$\frac{N_B}{\gamma}(\varepsilon_{t_0} - \varepsilon_{t_C}) > K$$

From which:

$$(\varepsilon_{t_0} - \varepsilon_{t_C}) > \frac{\gamma}{N_B}K$$

Lin *et al.* (2006) proposed the following close trade condition, *CTC (φ, ω),* in order to obtain the required minimum profit *K* on any completed trade: a position is opened under *OTC(φ)* by buying $N_B > \frac{K\gamma}{(\varphi-\omega)}$ shares of stock *B*, with $\varphi > \omega$, and selling $N_A = {N_B}/{\gamma}$ shares of stock *A*, then the position is closed at time t_C when $\varepsilon_{t_C} < \omega$. It is easy to demonstrate, based on equation (3.9), that $N_B > \frac{K\gamma}{(\varphi-\omega)}$ represents a sufficient condition to obtain a profit larger than *K* if the position opened under *OTC(φ)*, i.e. when $\varepsilon_{t_0} > \varphi$, is closed when $\varepsilon_{t_C} < \omega$. Consider the following example:

$$N_B = \frac{K\gamma}{(\varphi - \omega)} + 1$$

Recalling equation (3.13), the profits at time t_C can be rewritten as follows:

$$\pi_{t_C} = \frac{N_B}{\gamma}(\varepsilon_{t_0} - \varepsilon_{t_C}) = \frac{1}{\gamma}\left(\frac{K\gamma}{(\varphi - \omega)} + 1\right)(\varepsilon_{t_0} - \varepsilon_{t_C}) = \frac{(\varepsilon_{t_0} - \varepsilon_{t_C})}{(\varphi - \omega)}K + \frac{1}{\gamma}(\varepsilon_{t_0} - \varepsilon_{t_C})$$

Since $(\varepsilon_{t_0} - \varepsilon_{t_C}) > (\varphi - \omega)$ and $\frac{1}{\gamma}(\varepsilon_{t_0} - \varepsilon_{t_C}) > 0$, it is possible to conclude that:

$$\pi_{t_C} = \frac{(\varepsilon_{t_0} - \varepsilon_{t_C})}{(\varphi - \omega)} K + \frac{1}{\gamma}(\varepsilon_{t_0} - \varepsilon_{t_C}) > K$$

Applications of the CCW strategy

The authors test their approach using sample price data generated from the following cointegration model:

$$p_{A,t} - \gamma p_{B,t} = \mu + \varepsilon_t$$

where μ represents the equilibrium value, $\{\varepsilon_t\}$ is a stationary process and γ is the cointegration coefficient. Lin *et al.* (2006) demonstrate that under the *CCW* rule, the number of trades completed in a trading horizon is heavily affected by the open and close criterion. They demonstrate that the lower is the difference between the value of opening condition (φ) and the value of the closing condition (ω), the higher is the average total trade numbers in a certain trading period.

In addition, they also considered another simulation study, in which the total dollar investment permitted per trade is constrained. In other words, trades that require an investment above a certain threshold are not considered. The results demonstrate that the size of the average dollar commitment per trade required to meet the minimum profit per trade condition can make the rate of return on investment very small.

Finally, the authors test their strategy using the daily closing prices from January 2, 2001 to August 30, 2002 (20 months) for two Australian quoted banks: Australia New Zealand Bank and Adelaide Bank. Three results of this empirical application are remarkable:

- At least one valid trade is generated at all *MPPT* levels unless the open condition is too low to allow potential trades to develop at the given investment levels.
- The higher the size of the dollar commitment per trade, the higher the number of valid trades for a given *MPPT*.
- In general, a reduction of the open trade boundary value increases the number of valid trades, but it falls to zero trades when the spread becomes too small to generate trades within the investment level considered.

Despite the interesting results, the strategy developed by Lin *et al.* (2006) has several weaknesses. First, the *MPPT* is set in absolute terms, so the profitability scaled by initial investment can be very small. Second, the simulation study lacks diversity since it considers only one cointegration model. Finally, the empirical analysis only examines two shares over a very limited sample period of less than two year (Krauss, 2017).

3.5.2 Pairs Trading and Selection Methods: Is cointegration Superior? (Huck and Afawubo, 2015)

In their research Huck and Afawubo (2015) considers three different selection methods for pairs trading, specifically:

- *The distance approach*: each stock is paired with the 'matching partner' that minimizes the sum of Euclidean squared distances (*SSD*) and during the trading period a long/short position in a pair is opened whenever its normalized price difference diverge more than a prespecified trigger (see Section 1.2).
- *The stationarity of the price ratio approach*: it is based on the idea that in order to generate profits in a pair-trade, the price ratio between two stocks needs to have a constant mean and a constant volatility over time. In other words, the price ratio has to be a stationary process (see Section 2.2), so that any deviation from the equilibrium has a diminishing effect over time and finally disappears bringing the series back to its equilibrium value. In order to find price ratio with a constant mean and volatility, Huck and Afawubo (2015) use the Augmented Dickey Fuller test (see Section 2.7.2) and then select for the trading period only the pairs with the lowest *ADF t*-statistics.
- *The cointegration approach*: in this study potential cointegration between the stocks is examined using the Johansen test (see Section 2.11.2) and then cointegrated pairs with the highest trace statistics will be kept as eligible pairs for the trading period.

The objective of the study by Huck and Afawubo (2015) is to establish whether or not, one of these approaches can be considered superior, relative to the others, in terms of significant monthly positive returns.

The data used are the prices of the *S&P 500* stocks in the period from August 2000 to September 2011 (134 months), which are among the most liquid in the world, implying relatively low transaction costs. The authors consider two formation periods (one year (252 trading days) and two years (504 trading days)) and two opening triggers (2-standard deviation and 3-standard deviation) for each selection method, meaning that they perform four different specifications for each approach. It is important to notice that the greater the opening trigger, the lower the number of trades during the trading period, so the 3-SD rule is a more selective scheme with respect to the classical 2-SD rule used by most authors, such as Gatev *et al.* (2006) and Do and Faff (2010). A selection procedure starts every month (21 trading days) and the trading period of a pair lasts six months (126 trading days).

Returns Computation

The returns are computed on a daily basis as the mean return among all pairs (which must be at least ten) opened a given day in the entire portfolio, i.e. the sum of the six portfolios that start one month apart. It is possible that some days the whole portfolio is composed of less than ten pairs (especially if the more selective scheme is considered). In this case, the missing positions will be filled by a long position in the market index. The daily return (excluding transaction costs) of the portfolio at time *t* can be computed as follow:

$$R_{Portfolio,t} = \frac{1}{N_t}\sum_{i=1}^{N_t} R_{i,t}$$

where N_t is the number of open pairs at time *t*, $R_{i,t}$ is the return of the price ratio of the i^{th} pair at time *t*. If $N_t < 10$, $R_{i,t} = Y_t, \forall i \in [N_t + 1;\ 10]$, where Y_t is the return of the market at time *t*. The monthly transaction costs are estimated based on the three components identified by Do and Faff (2012): commissions (one way cost of 0.1%), market impact (one-way cost of 0.2%) and short-selling constraints (1% p.a. payable over the duration of each trade).

Empirical Results

Huck and Afawubo (2015) demonstrate that the strategies considered are based on very different groups of eligible pairs. Indeed, the percentage of identical eligible pairs among strategies is never above 13%. In other words, the proportion of pairs which are selected the same month by two different strategies is always below 13%.

The most important results obtained by the researchers are reported in *Table 3.1*, which considers the six specifications obtained considering a formation period of one year, and in *Table 3.2,* which reports the parametrizations obtained with a training period of two years.

Table 3.1: Result with 1-year formation period

	Design of the strategy (1-year formation period)					
Method	Distance		Stationarity		Cointegration	
Opening Trigger (nb of σ)	2	3	2	3	2	3
Monthly Returns						
(without transaction costs)	0,33	0,27	0,48	0,36	2,08	5,86
Consistent *p*-values (Hansen)	0,00	0,03	0,02	0,01	0,00	0,00
Monthly transaction costs	0,38	0,28	0,40	0,30	0,33	0,20
Monthly Returns						
(with transaction costs)	-0,05	-0,01	0,08	0,06	1,75	5,66
Consistent *p*-values (Hansen)	1,00	0,47	0,31	0,26	0,00	0,00
Trading statistics and portfolio composition (per pair, per 6-month period)						
Non-traded pairs (%)	4,93	20,75	5,07	14,93	3,81	43,21
Non convergent (NC) pairs (%)	45,49	52,95	41,68	53,40	60,00	44,66
Single round trip pairs (%)	33,28	51,38	34,18	26,16	28,28	11,87
Multiple opening pairs (%)	16,31	4,93	19,07	5,52	7,91	0,26
Profitable trades (%)	62,59	58,31	61,90	57,45	66,82	76,00
NC profitable trades (%)	27,93	37,85	26,11	35,41	47,76	69,79
NC unprofitable trades (%)	72,07	62,15	73,89	64,59	52,24	30,21

Source: N. Huck, K. Afawubo (2015)

Table 3.2: Result with 2-year formation period

	Design of the strategy (2-year formation period)					
Method	Distance		Stationarity		Cointegration	
Opening Trigger (nb of σ)	2	3	2	3	2	3
Monthly Returns						
(without transaction costs)	0,44	0,47	0,58	0,64	1,68	3,77
Consistent *p*-values (Hansen)	0,00	0,00	0,00	0,01	0,00	0,00
Monthly transaction costs	0,29	0,20	0,31	0,23	0,30	0,19
Monthly Returns						
(with transaction costs)	0,15	0,27	0,27	0,41	1,38	3,58
Consistent *p*-values (Hansen)	0,11	0,02	0,09	0,05	0,00	0,00
Trading statistics and portfolio composition (per pair, per 6-month period)						
Non-traded pairs (%)	19,03	45,86	13,43	34,89	12,09	47,16
Non convergent (NC) pairs (%)	50,60	41,64	51,42	49,40	56,53	43,21
Single round trip pairs (%)	24,25	10,82	27,54	13,62	25,97	9,14
Multiple opening pairs (%)	6,12	1,68	7,61	2,09	5,41	0,49
Profitable trades (%)	61,04	58,79	58,89	59,14	64,00	71,11
NC profitable trades (%)	38,40	45,78	35,17	46,00	45,94	64,89
NC unprofitable trades (%)	61,60	54,22	64,83	54,00	54,06	35,11

Source: *N. Huck, K. Afawubo (2015)*

The first result is that the monthly returns (ignoring transaction costs) are positive and statistically significant at 5% level for all the approaches, regardless of the length of the formation period and the opening trigger.

However, the strategies with monthly returns robust to transaction costs are the following:

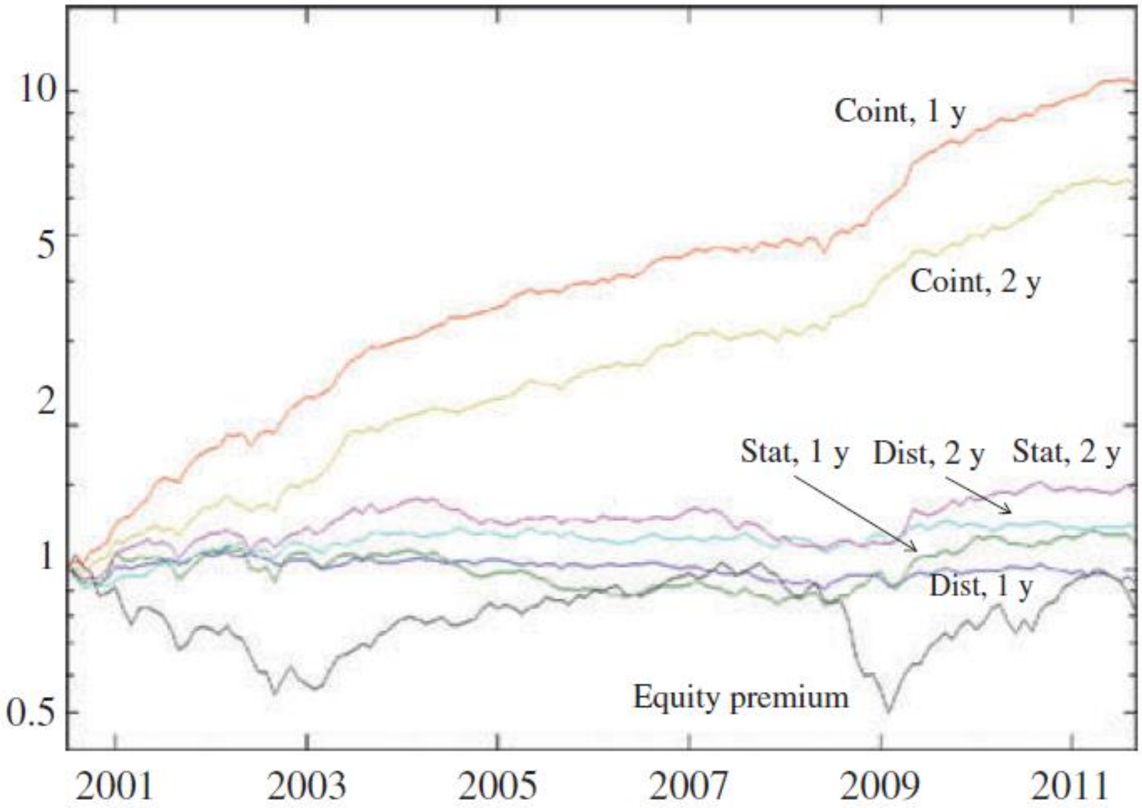
- Cointegration approach for all the parametrizations considered. With monthly returns ranging from 1.38% (2-year formation period and 2-SD rule) to 5.66% (1-year formation period and 3-SD rule).
- Distance approach with 2-year formation period and 3-SD as opening trigger (Monthly returns after transaction costs of 0.27%).
- Stationarity approach with 2-year formation period and 3-SD as opening trigger (Monthly returns after transaction costs of 0.41%).

The percentage of profitable trades is crucial for the success of pairs trading. In the experiments considered, this percentage ranges from 57.45% (stationarity approach with 1-year formation period and 3-SD rule) to 76% (cointegration approach with 1-year formation period and 3-SD rule). Despite these proportions are quite high, some strategies have weak and insignificant monthly returns. This can be explained by looking at the percentage of unprofitable nonconvergent trades that for the distance approach and the stationary approach are particularly high. It is not surprising that the most successful parametrizations, that are:

- Cointegration approach with 1-year formation period and 3-SD as trigger (Monthly returns (with transaction costs) of 5.66%)
- Cointegration approach with 2-year formation period and 3-SD as trigger (Monthly returns (with transaction costs) of 3.58%)

are the only strategies in which the percentage of profitable non-convergent trades stays above 50% (69.79% and 64.89%, respectively). As argued by Huck and Afawubo (2015), the cointegration approach seems to significantly reduce nonconvergent risk, or better the risk of losses in case of non-convergent trades.

***Figure 3.1**: Cumulative returns including transaction costs pairs trading strategies with a 2-SD trigger versus equity premium*



***Source:** N. Huck, K. Afawubo (2015)*

Finally, consider *Figure 3.1* which compares the cumulative returns (including transaction costs) of the different strategies with a 2-SD opening trigger versus the return of the market index (*S&P 500* index). Notice that, over the eleven years considered in this research, the cointegration-based pairs trading strategy significantly outperforms the alternative pairs trading strategies and also the market index. Therefore, it is possible to conclude that over the 134 months examined by Huck and Afawubo (2015) the pairs trading cointegrated strategy is clearly superior, in terms of monthly returns, relative to the other strategies. The main driver of the success of the cointegration-based pairs trading strategy seems to be its ability to identify econometrically more stable relationships compared to other approaches, which translates into a higher percentage of profitable non-convergent trades.

3.5.3 Markowitz Portfolio Optimization through Pairs Trading Cointegrated Strategy in Long-Term Investment (Naccarato *et al.*, 2019)

The objective of this research is to solve the Markowitz portfolio optimization problem (Markowitz, 1952) for a long-term horizon investment, through the pairs trading cointegrated strategy. In order to solve this problem, it is necessary to know the return and the risk of each stock included in the portfolio. The return of a stock depends on its price fluctuation over time and, therefore, its determination requires an estimation procedure. Naccarato *et al.* (2019) suggest estimating stock prices through a cointegration-based pairs trading strategy, which allows to identify the prices of each stock based on a cointegration relationship estimated with the Vector Error Correction Model (*VECM*) (see Section 2.10.4). The value of the estimated cointegration relationship represents the equilibrium between the prices, and the return of the stocks depends on the extent to which the two prices fluctuate around this equilibrium.

Since the authors aim at solving the Markowitz optimization problem using cointegration-based pairs trading strategy, this implies that the portfolio considered must contain pairs of cointegrated stocks. In particular, they consider the 30 stocks with the highest capitalization among all real European stock in the financial sector over the period from August 2008 to August 2018. Moreover, the authors consider stocks belonging to the same sector because, as discussed in Section 3.2, they are more likely to be cointegrated since they are affected by the same common factors. Among the 435 potential pairs that can be obtained combining the 30 selected stocks, Naccarato *et al.* (2019) find out, using the Johansen test (see Section 2.11.2), that only 148 pairs are actually cointegrated. Finally, the authors randomly select three pairs among the 148 cointegrated pairs for the implementation of their strategy.

Consider the i^{th} pair of cointegrated series (i = 1, 2, 3) consisting of stocks *A* and *B*, where $p_{A,t}$ and $p_{B,t}$ are the observed prices of stock *A* and *B* at time *t*, respectively. Their return can be computed as:

Stock A	$r_{iA,t} = log\ (p_{iA,t}) - log\ (p_{iA,t-1})$
Stock B	$r_{iB,t} = log\ (p_{iB,t}) - log\ (p_{iB,t-1})$

If the two time series $log\ (p_{iA,t})$ and $log\ (p_{iB,t})$ are cointegrated, this means that there exists a linear combination which is stationary, that is:

$$w_{i,t} = (log\ (p_{iA,t}) - \beta_i log\ (p_{iB,t})) \sim I(0)$$

where $\beta_i = (1, \beta_i)'$ is the vector of the cointegration coefficients for the pair *i*.

The time series $w_{i,t}$ fluctuates around an equilibrium relationship between the time series $log(p_{iA,t})$ and $log\ (p_{iB,t})$, which is represented by $\mu_{i,w} = E(w_{i,t})$.

The *VECM* for the two series $log\ (p_{iA,t})$ and $log\ (p_{iB,t})$ can be written as:

$$\begin{bmatrix} log(p_{iA,t}) - log(p_{iA,t-1}) \\ log(p_{iB,t}) - log(p_{iB,t-1}) \end{bmatrix} = \begin{bmatrix} \alpha_{i,A} \\ \alpha_{i,B} \end{bmatrix} (w_{i,t-1} - \mu_{i,w}) + \qquad (3.14)$$

$$+ \begin{bmatrix} \phi_{i,AA} & \phi_{i,AB} \\ \phi_{i,BA} & \phi_{i,BB} \end{bmatrix} \begin{bmatrix} log(p_{iA,t-1}) - log(p_{iA,t-2}) \\ log(p_{iB,t-1}) - log(p_{iB,t-2}) \end{bmatrix} + \begin{bmatrix} \varepsilon_{iA,t} \\ \varepsilon_{iB,t} \end{bmatrix}$$

where $\alpha_i = (\alpha_{i,A},\ \alpha_{i,B})'$ is the vector of the adjustment coefficients, $\varepsilon_{i,t} = (\varepsilon_{Ai,t}, \varepsilon_{iB,t})'$ represents the vector of disturbances, $\Gamma_i = \begin{bmatrix} \phi_{i,AA} & \phi_{i,AB} \\ \phi_{i,BA} & \phi_{i,BB} \end{bmatrix}$ is the 2×2 matrix of autoregressive coefficients and the term $(w_{i,t-1} - \mu_{i,w})$ represents the deviation from the equilibrium value of the two time series $log\ (p_{iA,t})$ and $log\ (p_{iB,t})$. Equation (3.14) describes the returns of the stocks *A* and *B* at time *t* as a function of their return at time *t-1* and of the deviation of $log(p_{iA,t})$ and $log\ (p_{iB,t})$ from the equilibrium relationship at time *t-1*.

The authors estimate the returns of stock *A* and *B* using the Engle-Granger methodology (see Section 2.11.1), and then they apply the pairs trading strategy to calculate the return of the i^{th} pair. The pairs trading strategy requires two steps:

- If at time *t* the trader notices a sufficient deviation below the equilibrium relationship, i.e. $w_{i,t} = \mu_{i,w} - \Delta_i$, the trader will buy one share of stock *A* and sell β_i shares of stock *B*.
- If at time $t + h_i$ the trader observes a sufficient deviation above the equilibrium value, i.e. $w_{i,t} = \mu_{i,w} + \Delta_i$, he/she will buy β_i shares of stock *B* and sell one share of stock *A*.

The pairs trading strategy is concluded at time $t + h_i$, when the double deviation from the equilibrium occurs, and the trader will obtain a return ($r_{i,t+h_i}$) equal to $2\Delta_i$.

Each pair of stocks produces the expected return ($r_{i,t+h_i}$) after a certain period h_i, which is generally different for each pair. Thus, the investment period of the Markowitz portfolio will correspond to that of the pair that takes the longest period of time to achieve the expected return. Once the return $r_{i,t+h_i}$ have been obtained, the next step concerns the resolution of the following problem of risk minimization, in order to find the allocation coefficients (θ_i) for each pair:

$$
\begin{cases}
min \frac{1}{2} \sum_{i,i'=1}^{3} \theta_i \theta_{i'} \sigma_{ii'} \\
\sum_{i=1}^{3} \theta_i r_{i,t+h_i} = \bar{r} \\
\sum_{i=1}^{3} \theta_i = 1
\end{cases}
\tag{3.15}
$$

where $\sigma_{ii'}$ is the covariance between the returns of the pair i and i', θ_i is the proportion of capital to be invested in each pairs of stocks and $\bar{r}$ is the expected overall return of the Markowitz portfolio. By varying the value of $\bar{r}$ in the interval [min ($r_{i,t+h_i}$), max ($r_{i,t+h_i}$)], with $i = 1, 2, 3$, the efficient frontier can be constructed.

However, the trader needs to know the number of shares to be purchased for each of the 6 stocks in the portfolio and not the proportion of capital to be invested in each pair. The proposal by Naccarato *et al.* (2019) to overcome this problem is based on the fact that the i^{th} product $\theta_i r_{i,t+h_i}$ in (3.15) is a function of the return of the single stocks:

$$\theta_i r_{i,t+h_i} = \theta_i r_{A,t+h_i} - \beta_i \theta_i r_{B,t+h_i}$$

Since the parameters that define the quantity $\theta_i r_{i,t+h_i}$ are the cointegration coefficient (β_i) and the capital to be invested in each pair (θ_i), it seems reasonable to obtain the allocation of the single stock ($\gamma_{i,A}$ and $\gamma_{i,B}$) from the linear combination of this two quantities, that is:

$$\begin{bmatrix} \gamma_{i,A} \\ \gamma_{i,B} \end{bmatrix} = \begin{bmatrix} \theta_i \\ -\beta_i \theta_i \end{bmatrix}$$

The last objective of the research by Naccarato *et al.* (2019) is to compare, by means of a bootstrap simulation, the results of the pairs trading cointegrated strategy (*PAIRS TRADING*) with five more methods, specifically the Autoregressive Integrated Moving Average model (*ARIMA*), Vector Autoregressive model (*VAR*), Capital Asset Pricing Model (*CAPM*), the Multifactor model (FACTOR), and the Dynamic Linear Model (*DLM*).

Multifactor models of asset returns can be divided into three types: *macroeconomic*, *fundamental*, and *statistical factor models*. All these models can be considered extensions to the CAPM model, which assumes that the returns of the assets are almost completely explicable by the behaviour of the overall market. Thus, the CAPM is based on a single explanatory factor and exposure value, the market return and *beta* (which measures the asset's linear sensitivity to the market), respectively (Vidyamurthy, p. 38). In contrast, the multifactor models consider multiple explanatory factors and exposure values in order to explain assets' returns, specifically (Connor, 1995):

- The *macroeconomic factor models* use historical asset returns and observable economic variables (e.g. inflation, percentage change in industrial production, excess return to long-term government bonds, etc.) as measures of the pervasive factors in asset returns. The main shortfall of these models is that they require identification and measurement of all the pervasive shocks affecting the asset returns, which is a very complex task (see Chen *et al.* (1986), Burmeister and McElroy (1988)).
- The *fundamental factor models* use company and industry attributes (e.g. firm size, dividend yield, book-to-market ratio, etc.) and market data as descriptors to explain the asset returns. The factors in a fundamental factor model are the realized returns to a set of mimicking portfolios designed to capture the marginal returns associated with a unit of exposure to each attribute. In other words, each factor is the realized return per extra unit of factor, holding other attributes constant (see Fama and French (1993) and Griffin (2002)).
- The *statistical factor models* use maximum-likelihood and principal-components based factor analysis procedures on cross sectional samples of asset returns to identify the pervasive factors in returns. The key advantage of these models is that the only information needed is the asses' prices, from which it is possible to calculate the returns which are used to estimate the statistical model. On the other hand, the main disadvantage of these models is that it could be difficult to provide an economic interpretation of the statistical factors (see Grinold *et al.* (1992)).

The *Dynamic Linear Model* (*DLM*), also called *Gaussian linear state-space model*, belongs to the class of state-space models. The state-space models are based on the idea that a time-series $\{Y_t\}$, denoting a vector of variables observed at date *t*, is an incomplete and noisy function of some underlying unobservable process $\{\theta_t\}$, called the state process (Petris *et al.*, 2009). Put differently, the observable process $\{Y_t\}$ depends on the latent process $\{\theta_t\}$, meaning that it is possible to assume that the observation process $\{Y_t\}$ only depends on the state of the system at the time the measurement is taken, $\{\theta_t\}$. The simplest *dynamic linear model* is the so-called *random walk plus noise* model, defined by the following system of equations (Petris *et al.* 2009):

Observation equation	$Y_t = \mu_t + v_t$	$v_t \sim N(0, V)$
State equation	$\mu_t = \mu_{t-1} + w_t$	$w_t \sim N(0, W)$

where the disturbances $\{v_t\}$ and $\{w_t\}$ are assumed to be uncorrelated at all lags, that is:

$$E(v_t, w_\tau') = 0 \qquad \textit{For all t and } \tau$$

In the model by Naccarato *et al.* (2019) the observations $\{Y_t\}$, i.e. the series of returns, are modelled as random fluctuations around a level $\{\mu_t\}$, which can evolve randomly over time (i.e. it is described by a random walk).

These five models are estimated using the same six time series of stocks considered for the cointegration-based pairs trading strategy, and each estimated model is used to solve the Markowitz problem, described in (3.15).

Ultimately, the six strategies are replicated 1000 times using the bootstrap methodology in order to compare their financial performance, which are summarized in *Table 3.3*.

Table 3.3: $\frac{\bar{r}}{s}$ *median and confidence intervals at 95% for the different strategies*

Method	Median	Lower Bound	Upper bound
PAIRS TRADING	24.9215	18.5500	34.3225
ARIMA	0.0335	0.0146	7.8548
CAPM	0.0196	0.0019	0.3975
FACTOR	0.0916	0.0007	0.3694
VAR	0.0037	0.0013	0.3951
DLM	0.1552	0.0015	0.3851

Sources: *A. Naccarato, A. Pierini, G. Ferraro (2019)*

The synthetic index used to compare the performance of the different models is the median of $\bar{r}/_{s}$, that is the ratio between the portfolio return and the corresponding minimum risk. The median of the *PAIRS TRADING* model is 24.9215, which is much higher than the other models for which this value is not higher than 0.1552 (*DLM*). Moreover, from the 95% confidence intervals it is possible to notice that the lower bound for the *PAIRS TRADING* model is significantly higher than the upper bounds of the confidence intervals for all the other methods. In other words, in 95% of cases, the *PAIRS TRADING* strategy provides a higher overall return per unit of risk relative to the other strategies.

Table 3.4: $\bar{r}$ *median values and confidence intervals at 95% for the different strategies*

Method	Median	Lower Bound	Upper bound
PAIRS TRADING	0.3050	0.2616	0.3451
ARIMA	0.0010	0.0010	0.2518
CAPM	0.0100	0.0001	0.0100
FACTOR	0.0100	0.0001	0.0100
VAR	0.0001	0.0001	0.0100
DLM	0.0100	0.0001	0.0100

Sources: *A. Naccarato, A. Pierini, G. Ferraro (2019)*

Table 3.5: *s median values and confidence intervals at 95% for the different strategies*

Method	Median	Lower Bound	Upper bound
PAIRS TRADING	0.0123	0.0090	0.0162
ARIMA	0.0346	0.0229	0.0684
CAPM	0.0367	0.0240	0.0537
FACTOR	0.0473	0.0255	0.1973
VAR	0.0364	0.0237	0.0856
DLM	0.0376	0.0241	0.0834

Sources: *A. Naccarato, A. Pierini, G. Ferraro (2019)*

The ratio $\bar{r}/_{s}$ can be very high for two reasons:

- A very high portfolio return ($\bar{r}$) for a given level of risk (s).
- A very low level of risk (s) for a given level of portfolio return ($\bar{r}$).

Thus, it can be useful to evaluate $\bar{r}$ and s separately for all the different methods. The results are reported in *Table 3.4* and *Table 3.5*.

The median value of the portfolio's total return is much higher than those obtained with the other models; in 50% of the cases the *PAIRS TRADING* model has a portfolio return value of 0.3050, which is thirty times the value of the best of the other models (0.0100). Moreover, the median of the risk *s* is much smaller compared with the risk associated with the other strategies. This means that the better results reported in *Table 3.3* are due to both higher return and lower risk.

It is important to highlight that the authors are comparing the cointegration-based pairs trading method, which is based on strategies of entry/exit rules in the different cointegrated pairs, to simple factorial models which do not imply any particular strategy, but simply represent different descriptions of the returns of the stocks considered. In other words, they do not demonstrate that the cointegration-based pairs trading approach is necessarily a better strategy than other types of dynamic trading, but they simply demonstrate that pairs trading strategy has significant returns compared to the market. In conclusion, their study shows that the cointegration-based pairs trading is profitable, but the other five models under analysis cannot be considered as competing strategies.

Chapter 4

Empirical Analysis

This chapter is dedicated to our empirical analysis. In particular, we explore the performance of different cointegration-based pairs trading strategies, using the daily closing stock prices of the major banks in the Italian banking system over the period from 2 January 2015 to 30 December 2019. *Section 4.1* presents in detail the data that we have used to assess the profitability of our pairs trading strategies. *Section 4.2* describes the first phase of the cointegration analysis which consists in the verification of the order of integration of the time series under analysis. In our work, this preliminary analysis is carried out using the Augmented Dickey-Fuller test (see *Section 2.7.2*) and the KPSS test (see *Section 2.8*). *Section 4.3* describes the trading design that we have selected for our pairs trading strategies, which is based on the work by Huck and Afawubo (2015). In particular, this study considers two different lengths for the formation period (1 year and 2 years) and one length for the trading period (6 months) with two opening triggers, which are based on a standard deviation metric (2-standard deviations and 3-standard deviations). Moreover, we decide to widen the pairs trading strategy proposed by Huck and Afawubo (2015) considering two different closing triggers, which are the convergence of the spread to its long-term equilibrium and the re-convergence of the spread to the opening trigger (i.e. the first time that the spread crosses the opening trigger the position is opened and the second time the position is unwound). Finally, in *Section 4.4* we present the results of our empirical analysis for the different parametrizations considered.

4.1 Data

For the purpose of our empirical analysis we consider four out of five major listed Italian banks, as defined in the research conducted by Mediobanca (2019). The reasons behind the choice of selecting only four out of five banks are explained below in this section. The ranking is based on the level of total tangible assets as of 31 December 2018. The resulting classification of this study is reported in *Table 4.1*.

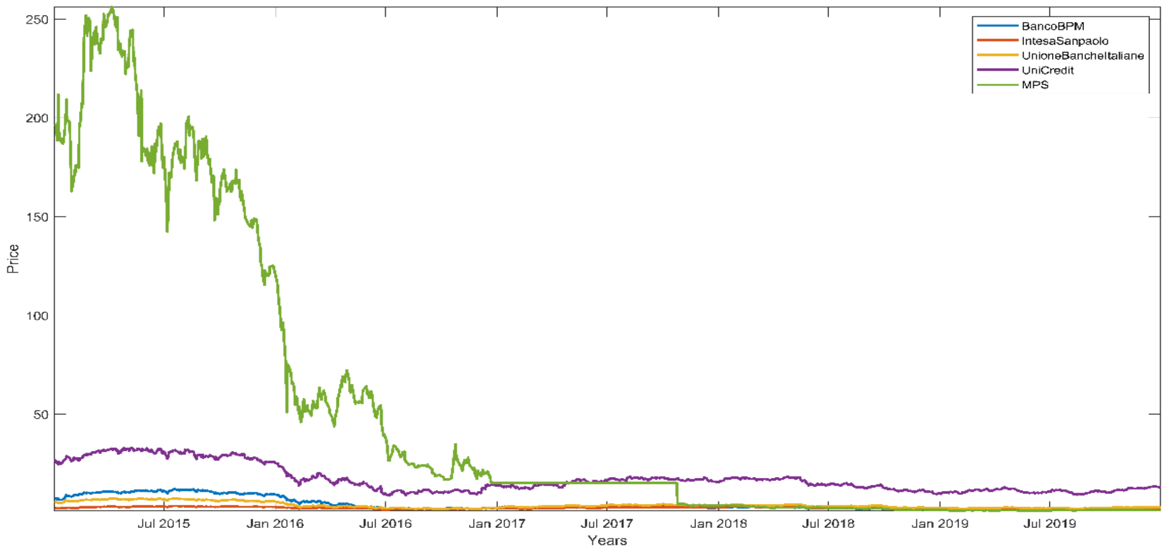
Table 4.1: *List of the major Italian banks by total tangible assets (data as of December 2018)*

Ranking	*Banks*	*Total Tangible Assets (Thousands €)*
1	UniCredit S.p.A.	827,961,000
2	Intesa Sanpaolo S.p.A.	778,624,000
3	Banco BPM S.p.A.	159,186,850
4	Banca Monte dei Paschi di Siena S.p.A.	130,267,082
5	Unione Banche Italiane S.p.A.	123,576,082

Source: *Mediobanca (2019).*

The data used in our study, retrieved on 14 May 2020 from the Thomson Reuters Eikon Database, consist in Euro-denominated daily closing stock prices of the five aforementioned banks. As for the time period, a 5-year interval from 2 January 2015 to 30 December 2019 provides a total of 1268 daily observations for each financial institution. *Figure 4.1* provides a graphical representation of our time series.

Figure 4.1*: Daily prices for Banco BPM, Intesa Sanpaolo, UniCredit, Unione di Banche Italiane, Banca Monte dei Paschi di Siena.*



We have decided not to consider Banca Monte dei Paschi di Siena for two main reasons: first, because of the significant fluctuations experienced by the stock price over the last five years compared to the other competitors under analysis. From a simple look at *Figure 4.1* it is clear that there were exceptional events in the time series of Monte dei Paschi di Siena over the last five years, which are also confirmed by the results summarized in *Table 4.2*. Second, due to a lack of data during the period between 23 December 2016 and 24 October 2017 (for a total of 213 daily observations). During this period of time, there was a temporary suspension of trading, ordered by CONSOB[4], on regulated markets, multilateral trading systems and Italian systematic internalization systems in relation to securities issued or guaranteed by Banca Monte dei Paschi and to financial instruments having securities issued by Banca Monte dei Paschi standing as underlying.

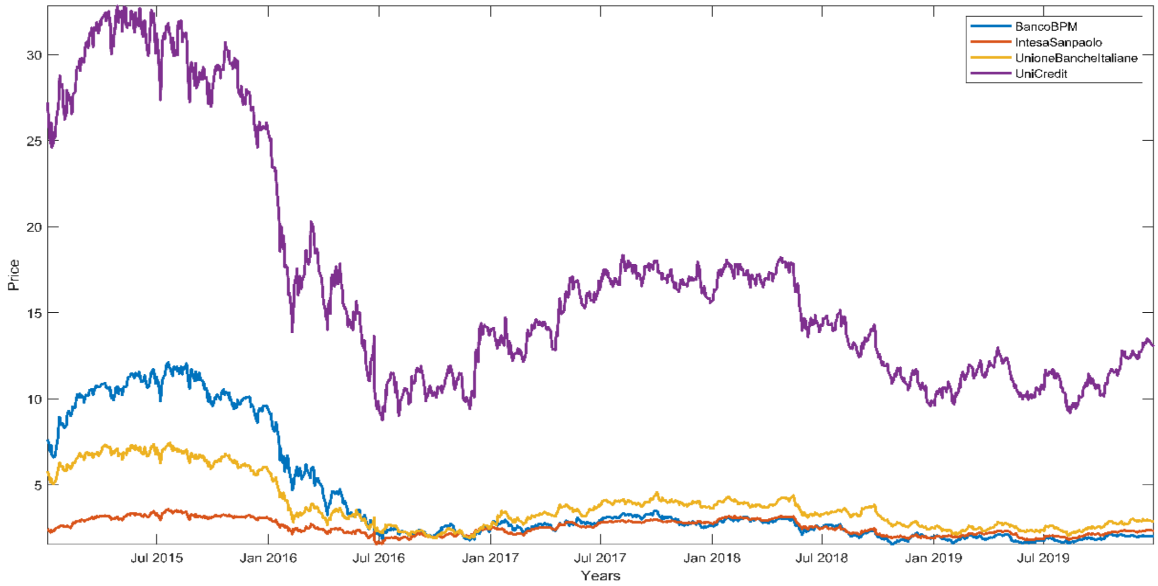
[4] CONSOB (Commissione Nazionale per le Società e la Borsa) is the Italian government authority responsible for regulating the Italian securities market.

Thus, the financial institutions considered for the purpose of this study are the following:

- UniCredit S.p.A.
- Intesa Sanpaolo S.p.A.
- Banco BPM S.p.A.
- Unione di Banche Italiane S.p.A. (Ubi)

Figure 4.2 shows a graphical representation of the time series of these four banks.

***Figure 4.2**: Daily prices for Banco BPM, Intesa Sanpaolo, UniCredit and Ubi.*



From the results reported in *Table 4.2*, we can obtain some insights regarding the situation of the five credit institutions considered over the period from 2 January 2015 to 30 December 2019:

- The average stock price of Banco BPM is 4.22 €, ranging from a minimum price of 1.56€ (registered on 26 October 2018) to a maximum price of 12.12€ (registered on 20 July 2015), with a standard deviation of 3.24.
- The average stock price of Intesa Sanpaolo is 2.54 €, ranging from a minimum price of 1.55€ (registered on 27 June 2016) to a maximum price of 3.60€ (registered on 20 July 2015), with the lowest standard deviation among the banks considered (0.46).
- The average stock price of Ubi is 3.82 €, ranging from a minimum price of 1.90€ (registered on 29 September 2016) to a maximum price of 7.44€ (registered on 23 July 2015), with a standard deviation of 1.53.
- The average stock price of UniCredit is 16.98 €, ranging from a minimum price of 8.78€ (registered on 7 July 2016) to a maximum price of 32.82€ (registered on 27 April 2015),

with the highest standard deviation among the four banks selected for our analysis (6.87).

- The average stock price of Banca Monte dei Paschi di Siena is 49.52 €, this value is obtained by keeping the stock price constant to 15.08€ (quoted price as of 22 December 2016) during the period from 23 December 2016 to 24 October 2017, ranging from a minimum price of 1.00 € (registered on 6 June 2019) to a maximum price of 256.16€ (registered on 7 April 2015), with a standard deviation of 73.25, which is more than 10 times the standard deviation of UniCredit.

Table 4.2: *Summary statistics: January 2015-December 2019*

Banks	*Mean*	*Median*	*St Dev*	*Min*	*Max*	*Skew*	*Kurt*
Banco BPM	4.22	2.72	3.24	1.56	12.12	1.33	3.08
Intesa Sanpaolo	2.54	2.46	0.46	1.55	3.60	0.23	1.86
Unione Banche Italiane	3.82	3.37	1.53	1.90	7.44	1.01	2.78
UniCredit	16.98	14.71	6.87	8.78	32.82	1.07	2.82
Banca MPS	49.52	15.08	73.25	1.00	256.16	1.48	3.69

4.2 Preliminary Analysis: Testing for the Order of (Co-) Integration

The first phase of a cointegration analysis consists in the verification of the order of integration of the individual time series considered. As discussed in Section 2.10.1, two time series, both integrated of order one, are said to be cointegrated if there exists a non-zero linear combination that is stationary, i.e. integrated of order zero. Therefore, the purpose of our preliminary analysis is to test whether the time series that we have selected are effectively integrated of order one. This can be done through unit root tests (see Section 2.7) and the stationarity tests (see Section 2.8). As highlighted in Section 2.8, the stationarity tests have been developed to resolve the low test power of the unit root tests: a time series with unit root close to one that was typically found non-stationary with the unit root tests can be correctly found stationary with the stationarity test. Although the stationarity tests resolve the low test power of the unit root tests, it is important to be aware that the results obtained with statistical testing are probabilistic, and so there is always a non-zero chance of being wrong. In order to increase the probability of a right inference, a clever approach that should be used in unit root testing is to combine stationarity tests and unit root tests. In particular, in this work we decide to use one unit root test, specifically the Augmented Dickey-Fuller test (see Section 2.7.2) together with one stationarity test, specifically the KPSS test (see Section 2.8).

We recall that the unit root tests are statistical tests which evaluate the null hypothesis that a generic time series $\{X_t\}$ is non-stationary against the alternative hypothesis of stationarity, that is:

$$H_0: X_t \sim I(1) \qquad vs. \qquad H_1: X_t \sim I(0)$$

Conversely, the stationarity tests evaluate the null hypothesis that a generic time series $\{X_t\}$ is stationary against the alternative hypothesis of no stationarity, that is:

$$H_0: X_t \sim I(0) \qquad vs. \qquad H_1: X_t \sim I(1)$$

In our study, we divide the verification procedure of the time series' order of integration in two steps:

1. First, we perform the aforementioned tests on each time series considered (Banco BPM, Intesa Sanpaolo, Ubi and UniCredit) in order to verify whether the data are stationary or non-stationary. The results of the Augmented Dickey-Fuller test are reported in *Table 4.3*, while those of the KPSS test are shown in *Table 4.4*.

Table 4.3: *Augmented Dickey-Fuller test results*

	BPM	Intesa	Ubi	UniCredit
Reject the Null Hypothesis	No	No	No	No
ADF test (stat. value)	-1.0322	-1.9902	-1.4930	-1.5664
ADF Critical Value (5%)	-2.8647	-2.8647	-2.86417	-2.8647
ADF test (p-Value)	0.7251	0.3009	0.5211	0.4886

The *p*-Values are left-tail probabilities. In MATLAB when the test statistics are outside the tabulated critical values, the function *adftest* returns maximum *p*-Values of 0.999 or minimum *p*-Values of 0.001.

Table 4.4: *KPSS test results*

	BPM	Intesa	Ubi	UniCredit
Reject the Null Hypothesis	Yes	Yes	Yes	Yes
KPSS test (stat. value)	83.1467	35.245	63.0604	71.0604
KPSS Critical Value (5%)	0.4630	0.4630	0.4630	0.4630
KPSS test (p-Value)	< 0.0100	< 0.0100	< 0.0100	< 0.0100

The *p*-Values are right-tail probabilities. In MATLAB when the test statistics are outside the tabulated critical values, the function *kpsstest* returns maximum *p*-Values of 0.100 or minimum *p*-Values of 0.01.

From the results of the Augmented Dickey-Fuller test it is possible to ascertain that all the four variables fail to reject the null hypothesis of non-stationarity at a 5% confidence level. The consistency of these results is supported by the values obtained for the KPSS test, whose values indicate that the tests strongly reject the null hypothesis of stationarity, for all the variables under analysis, in favour of the alternative of non-stationarity.

2. The second step consists in the implementation of the same tests on the differenced series, to check whether these data show any evidence of unit roots. If the unit root tests reject the null hypothesis of non-stationarity, or if the stationarity tests do not reject the null hypothesis of stationarity, we can conclude that the differenced time series under analysis are stationary processes. In *Table 4.5* and *Table 4.6* are reported the results of the Augmented Dickey-Fuller test and of the KPSS test for the differenced data, respectively.

Table 4.5: *Augmented Dickey-Fuller test results: Differenced data*

	BPM	Intesa	Ubi	UniCredit
Reject the Null Hypothesis	Yes	Yes	Yes	Yes
ADF test (stat. value)	-34.0714	-36.5771	-35.4759	-37.0759
ADF Critical Value (5%)	-2.8647	-2.8647	-2.8647	-2.8647
ADF test (p-Value)	< 0.001	< 0.001	< 0.001	< 0.001

The *p*-Values are left-tail probabilities. In MATLAB when the test statistics are outside the tabulated critical values, the function *adftest* returns maximum *p*-Values of 0.999 or minimum *p*-Values of 0.001.

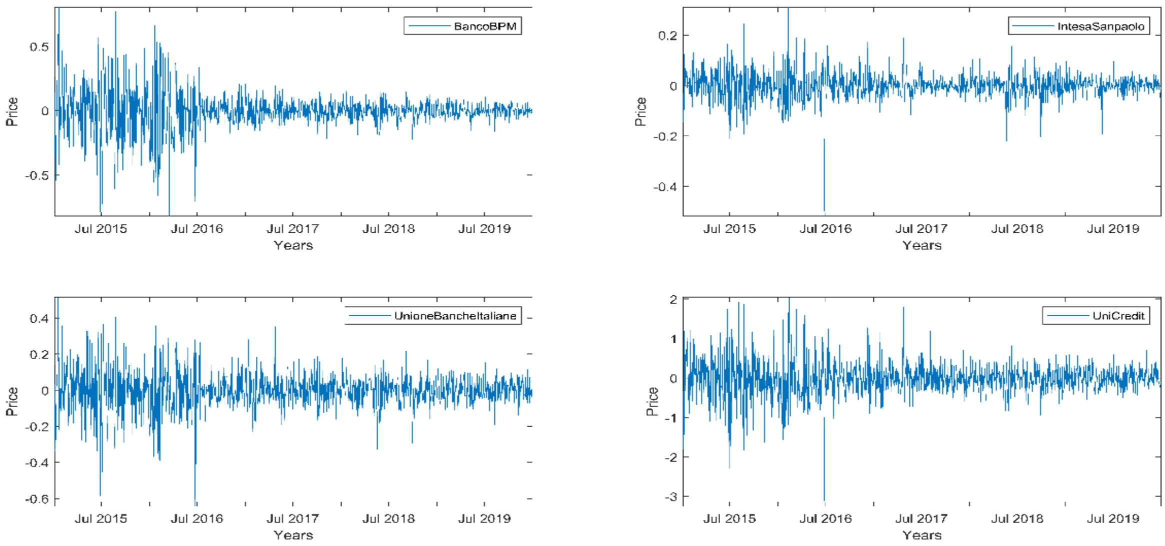
Table 4.6: *KPSS test results: Differenced data*

	BPM	Intesa	Ubi	UniCredit
Reject the Null Hypothesis	No	No	No	No
KPSS test (stat. value)	0.1793	0.0636	0.1111	0.1272
KPSS Critical Value (5%)	0.4630	0.4630	0.4630	0.4630
KPSS test (p-Value)	> 0.1000	> 0.1000	> 0.1000	> 0.1000

The *p*-Values are right-tail probabilities. In MATLAB when the test statistics are outside the tabulated critical values, the function *kpsstest* returns maximum *p*-Values of 0.100 or minimum *p*-Values of 0.01.

Focusing on the results of the ADF test: the four variables strongly reject the null hypothesis of presence of a unit root at a 5% confidence level, in favor of the alternative of stationarity. Also in this case, the results of the unit root tests are validated by the outcomes of the stationarity tests, which do not reject the null hypothesis of stationarity. Therefore, since there is strong evidence that the processes are nonstationary in levels (point 1), but they are stationary in first difference, it is possible to conclude that the variables under consideration are all integrated processes of order one. *Figure 4.3* provides a graphical representation of the differenced time series of Banco BPM, Intesa Sanpaolo, Ubi and UniCredit. From these figures it is possible to observe that the data differenced once seem to be effectively stationary processes, indeed they tend to oscillate along a stationary mean of about zero.

Figure 4.3: Differenced time series: Banco BPM, Intesa Sanpaolo, UniCredit and Ubi.



4.3 Methodology: Trading Design and Returns Computation

Following Huck and Afawubo (2015) (see Section 3.5.2), in this study we consider two different lengths for the formation (training) period, which represents the time interval during which the parameters of the experiment are computed (1 years and 2 years), and one length for the trading period, which represents the time period during which the experiment is run using the parameters computed in the training period (6 months), combined in the following two strategies:

1. Formation period: 1 year (252 days) and trading period: 6 months (126 days).
2. Formation period: 2 years (504 days) and trading period: 6 months (126 days).

Consider a generic formation period of M days, the idea is to test if two time series are cointegrated over this training period adopting the Engle-Granger approach (see Section 2.11.1). If evidence of cointegration exists then a pairs trading strategy will be implemented in the following N days (trading period) based of the cointegrating relationship found. Conversely, if the time series are found not to be cointegrated, no pairs trading strategy will be implemented in the subsequent trading period. In other words, we are using the information of the preceding M days to estimate the cointegrating relationship in the following N days. A new formation period is initiated after N days and the procedure is repeated in a rolling window fashion until the end of the sample considered.

As discussed in Section 3.4, whenever a cointegrating relationship is detected, in order to implement our pairs trading strategy, we have to select some trading rules to determine when to open and when to close a position. First, we have to calculate the spread between the two time series considered, e.g. stock A and stock B, over the subsequent trading period, which can be defined as the scaled difference of the two stocks' prices:

$$Spread_t = p_t^A - \gamma p_t^B$$

where γ represents the cointegration coefficient estimated during the formation period.

Accordingly, we compute the dimensionless *z-score* (or normalized spread), which can be calculated as:

$$z_t = \frac{Spread_t - \mu_e}{\sigma_e}$$

where μ_e is the spread's mean and σ_e is the spread's standard deviation, both calculated using the data of the formation period. Following Gatev *et al.* (2006), our trading signals for opening a position are based on a standard deviation metric. In particular, we consider two different triggers:

- 2-standard deviations (2-SDs)
- 3-standard deviations (3-SDs)

Consider a general q-standard deviation rule, whenever the normalized spread (z_t) hits the lower threshold, i.e. $z_t < -q$, this means that the portfolio of pairs considered is below its long-run equilibrium and so we should purchase it (long position), which means simultaneously buy one share of stock A and sell γ shares of stock B. Conversely, if the normalized spread hits the upper threshold, i.e. $z_t > q$, this means that the portfolio of pairs in overvalued and so we should sell it (short position), which means simultaneously sell one share of stock A and purchase γ shares of stock B.

Furthermore, we decide to consider two closing signals:

- Closing the position when: $z_t = 0$, which indicates that the normalized spread has reverted back to its long-run mean.
- Closing the position when the normalized spread crosses the opening trigger twice, which means that the first time that the *z-score* crosses the opening trigger a pairs trading position is opened and the second time the position is unwound. Analytically this can be stated as follows:

  Closing a short position when: $z_t < q$ with $q = 2{,}3$.

  Closing a long position when: $z_t > -q$ with $q = 2{,}3$.

In both cases, pairs that open but never converge to the closing threshold will only produce a cash flow on the last day of the trading period when all the positions will be close out. The cash flows generated by this type of trades can be either positive or negative depending on the spread between prices on the last day of the trading period. As explained in Section 3.4, the fact that price spreads may continue to diverge after position opening rather than revert to the long-run equilibrium value, due for example to significant pricing inefficiencies or particular market events, is one of the main risks associated to pairs trading strategies which can lead to substantial losses for the investor when he/she is forced to close his positions on the last day of the trading period.

As discussed in Section 3.4, it is not possible to establish a priori whether total profits increase or decrease with different opening triggers. In fact, the higher the opening threshold the lower the number of openings and trades during the trading period considered, but the higher the potential profits on each completed transaction. In contrast, a lower opening threshold will produce a higher number of trades during a specific trading period, which could potentially lead to higher total profits.

Moreover, it is important to notice that in our strategy the total profits also depend on the closing trigger selected. If we consider the case in which a position is closed whenever the normalized spread reverts back to its long-run mean, the number of completed trades during a particular trading period will be lower compared to the case in which the position is closed once the *z-score* cross the opening trigger twice, but the potential profits for each completed trade will be higher.

In this study we consider all the possible pairs that can be obtained with the four time series introduced in Section 4.2 ($\frac{N\times(N-1)}{2} = 6$). As discussed in Section 3.3, the Engle-Granger approach is not invariant or robust with respect to the direction of normalization, which means that the result of the estimated cointegrating relationship could change according to the choice of the dependent and independent variable, creating a potential problem of ambiguity (see Section 2.11.1). Since in the cases considered in this research it does not seem to exist a natural choice of dependent and independent variable, we decide to conduct for each pair under analysis both cointegration relationships with either one bank or the other as independent variable, that is:

$$p_t^A = \mu_1 + \gamma_1 p_t^B + \varepsilon_{1t}$$
$$p_t^B = \mu_2 + \gamma_2 p_t^A + \varepsilon_{2t}$$

where p_t^A and p_t^B represent the stocks' prices of the two banks considered, μ_1 (μ_2) and γ_1 (γ_2) represent the equilibrium value and the cointegration coefficient respectively resulting from the first (second) regression, and ε_{1t} and ε_{2t} represent the disturbance term in the equilibrium for the first and the second regression, respectively.

The cumulative profits of each strategy are calculated as the algebraic sum of the cash flows which occur during each trading period over the time horizon of 5 years considered, which can be:

- Positive: if a position is opened and unwound during a particular trading period.
- Positive or negative: if a position is opened but never converges to the closing threshold, and so we are forced to close it on the last day of the trading period.

Returns Computation

Consider two stocks A and B and their prices at a generic time *t*, which are respectively p_t^A and p_t^B, their spread at time *t* is defined as the scaled difference in the price of the two stocks, that is:

$$Spread_t = p_t^A - \gamma p_t^B \tag{4.1}$$

where γ is the cointegration coefficient. Let ϑ_t^A and ϑ_t^B be the number of units of stock A and stock B held in the portfolio at time *t*, respectively. Assuming a positive cointegration coefficient, that is $\gamma > 0$, a long (short) position on stock A and a short (long) position on stock B correspond to $\vartheta_t^A > 0$ ($\vartheta_t^A < 0$) and $\vartheta_t^B < 0$ ($\vartheta_t^B > 0$), respectively.

In order to exploit the cointegrating relationship between the two stocks under analysis, the following relation must be satisfied:

$$\vartheta_t^B = -\gamma \vartheta_t^A \tag{4.2}$$

Indeed, if relation (4.2) holds, then the value of the portfolio can be calculated as:

$$\vartheta_t^A p_t^A + \vartheta_t^B p_t^B = \vartheta_t^A (p_t^A - \gamma p_t^B) \tag{4.3}$$

from which it is possible to ascertain that the value of the portfolio at time *t* is proportional to the spread, as defined in equation (4.1).

Pairs trading relies on the idea that if prices diverge sufficiently from the long-term equilibrium value, there is a chance to make a profit because they are expected to converge to their equilibrium value sooner or later. Whenever a trading signal occurs, that is when the normalized spread hits one of the pre-specified opening threshold, the trade is opened by simultaneously buying (long position) the undervalued stock, whose price is expected to increase, and selling (short position) the overvalued stock, whose price is expected to decrease. Assuming that the money raised from shorting a stock can be immediately invested to buy the other stock, these positions are self-financing and do not require any capital to trade. The transaction is then closed by reverting the opening positions once the *z-score* hits the pre-determined closing threshold.

As discussed by Broussard and Vaihekoski (2012) the computation of the return on a zero net capital transaction is a problematic concept and no standardized method is available. In the literature different methodologies have been proposed to overcome this problem: returns can be computed either on the long leg of the position (i.e. on the cost incurred to buy the undervalued stock), the margin capital needed to undertake the short position, or on the gross capital exposure, intended as the sum of the long leg of the position and the absolute value of the short leg of the position.

In this study we decide to calculate the returns on the long leg of the position, assuming an initial capital of 1€. Our methodology relies on the assumption that the proceeds deriving from the short-selling of the overvalued stock are not immediately available to be invested in the long position, but they remain deposited in the broker's account until the end of the transaction. Consider the following strategy: at time t_1 the normalized spread z_{t_1} indicates that the portfolio composed by stock A and stock B is undervalued. The investor, at this point, should buy the undervalued stock (long position on stock A) and sell the overvalued stock (short position on B). Assuming that all the proceeds from the short position remain deposited in the broker's account, at time t_1 the investor with an initial wealth of 1€ can purchase $\vartheta_{t_1}^{A} = \frac{1}{p_{t_1}^{A}}$ shares of stock A. At time t_2 the *z-score* reverts back to its long-term equilibrium, so that a closing signal occurs; at this point the investor closes out his strategy by reverting the previously opened positions, and the profit or loss (r_{t_2}) can be computed as:

$$r_{t_2} = \vartheta_{t_1}^{A}(p_{t_2}^{A} - p_{t_1}^{A}) + \vartheta_{t_1}^{B}(p_{t_2}^{B} - p_{t_1}^{B}) = \vartheta_{t_1}^{A}\Delta p_{t_2}^{A} + \vartheta_{t_1}^{B}\Delta p_{t_2}^{B} \quad (4.4)$$

Recalling the relationship described in (4.2), the equation (4.4) can be rewritten as:

$$r_{t_2} = \vartheta_{t_1}^{A}\Delta p_{t_2}^{A} + \vartheta_{t_1}^{B}\Delta p_{t_2}^{B} = \frac{1}{p_{t_1}^{A}}(\Delta p_{t_2}^{A} - \gamma\Delta p_{t_2}^{B})$$

Now consider the case in which at time t_1 the normalized spread z_{t_1} indicates that the portfolio composed by stock A and stock B is overvalued. The investor should purchase the undervalued stock (long position on stock B) and sell the overvalued stock (short position on stock A). Assuming that all the revenues from the short position remain deposited in the broker's account, at time t_1 the investor with an initial capital of 1€ can purchase $\vartheta_{t_1}^{B} = \frac{1}{p_{t_1}^{B}}$ shares of stock B, while the number of shares of stock A which are sold short can be determined using equation (4.2). The return of the investor at the closing time t_2, i.e. when the *z-score* hits the pre-specified closing trigger, can be calculated as:

$$r_{t_2} = \vartheta_{t_1}^{A}\Delta p_{t_2}^{A} + \vartheta_{t_1}^{B}\Delta p_{t_2}^{B} = -\frac{1}{\gamma p_{t_1}^{B}}(\Delta p_{t_2}^{A} - \gamma\Delta p_{t_2}^{B})$$

Suppose that the strategy is re-opened at time t_3, at this date the available capital (W_{t_3}) of the investor will be equal to the initial capital, which was assumed to be 1€, plus the profits/losses generated by the previous transaction, that is:

$$W_{t_3} = 1 + \vartheta_{t_1}^A \Delta p_{t_2}^A + \vartheta_{t_1}^B \Delta p_{t_2}^B$$

This capital will be completely invested in the stock which will result to be undervalued at time t_3. In general, the available capital of the investor at time t_n (W_{t_n}) can be calculated as:

$$W_{t_n} = W_{t_{n-1}}(1 + \vartheta_{t_{n-1}}^A \Delta p_{t_n}^A + \vartheta_{t_{n-1}}^B \Delta p_{t_n}^B)$$

where $W_{t_{n-1}}$ is the available capital at time t_{n-1} and ($\vartheta_{t_{n-1}}^A \Delta p_{t_n}^A + \vartheta_{t_{n-1}}^B \Delta p_{t_n}^B$) represents the return at time t_n by the strategy initiated at time t_{n-1}.

Thus, in order to calculate the return of the entire strategy it is sufficient to subtract the initial capital of 1€ from the final available wealth, that is:

$$r_{t_n}^S = (W_{t_n} - 1)$$

4.4 Empirical Results

4.4.1 Closing Trigger: $z_t = 0$

In this section we present the results obtained in the case in which the pairs trading positions are closed when the *z-score* reverts back to its long-term equilibrium, i.e. when $z_t = 0$. The results reported in *Table 4.7* represent the annualized returns of the twelve pairs analysed, both considering a length of the formation period of one year (252 days) with trading period of 6 month (126 days), and a formation period of two years (504 days) with a trading period of 6 months (126 days), these parametrizations are analysed considering two opening triggers set at 2-standard deviations and 3-standard deviations, respectively.

Notice that four pairs, which are *Ubi_BPM, BPM_UniCredit*, *UniCredit_BPM* and *Intesa_Ubi*, result not to be cointegrated during the entire time horizon of 5 years considered. Excluding these four pairs from the analysis, we have a total of 40 annualized returns which can be divided as follows:

- 60% positive annualized returns (24 strategies out of 40).
- 32.5% negative annualized returns (13 strategies out of 40).
- 7.5 % zero annualized returns (3 strategies out of 40).

The positive annualized returns range from 9.94% (*UniCredit_BPM,* formation period of 2 year and trading period of 6 months with opening trigger set at 2-standard deviations) to 85.93% (*Intesa_BPM*, formation period of 1 year and trading period of 6 months with opening trigger set at 2-standard deviations); the negative returns range from -4.66% (*BPM_Intesa*, formation period of 2 years and trading period of 6 months with opening trigger set at 3-standard deviations) to -67.39% (*BPM_Intesa*, formation period of 1 year and trading period of 6 months with opening trigger set at 2-standard deviations). All the positive annualized returns are due to single or multiple profitable trades which are not offset by unprofitable trades deriving from the closure of the positions on the last day of the trading period before the convergence to the closing trigger has occurred. Conversely, the negative annualized returns are the result of single unprofitable trades which occur on the last day of the trading period when all the opened positions are closed out, i.e. positions that open during the trading period considered but never converge to the closing trigger, which completely offset the positive returns deriving from successfully completed trades, if any (see *Table B1* in the *Appendix B*).

The first thing that shall be noticed is that in most of the cases considered, when we increase the opening trigger from 2-standard deviations to 3-standard deviations, the annualized returns increase. This is due to the fact that opening a position when the *z-score* is less than or equal to -3, which corresponds to a deviation of the spread from the long-term equilibrium of 3-standard deviations, allow us to purchase the portfolio, which in this case is undervalued with respect to its long-term equilibrium, at a more convenient price compared to the case in which the same position is opened when the *z-score* is less than or equal to -2. Similarly, opening a position when the *z-score* is greater than or equal to +3, allow us to sell the portfolio, which in this case results to be overvalued with respect to its equilibrium value, at a more convenient price compared to the case in which the position is opened when the normalized spread is greater or equal to +2.

Table 4.7: *Annualized returns with closing trigger:* $z_t = 0$

Formation Period, Trading period	***Opening trigger***	***BPM_Intesa***	***Intesa_BPM***
1 year, 6 months	2 SDs	-67.39%	85.93%
	3 SDs	-28.74%	-10.42%
2 years, 6 months	2 SDs	-17.56%	-50.61%
	3 SDs	-4.66%	-23.39%
Formation Period, Trading period	***Opening trigger***	***BPM_Ubi***	***Ubi_BPM***
1 year, 6 months	2 SDs	37.54%	Not cointegrated
	3 SDs	0%	Not cointegrated
2 years, 6 months	2 SDs	-28.91%	-56.98%
	3 SDs	-8.19%	-34.92%
Formation Period, Trading period	***Opening trigger***	***BPM_UniCredit***	***UniCredit_BPM***
1 year, 6 months	2 SDs	Not cointegrated	Not cointegrated
	3 SDs	Not cointegrated	Not cointegrated

2 years, 6 months	2 SDs	18.24%	9.49%
	3 SDs	44.90%	31.47%
Formation Period, Trading period	***Opening trigger***	***Intesa_Ubi***	***Ubi_Intesa***
1 year, 6 months	2 SDs	Not cointegrated	24.29%
	3 SDs	Not cointegrated	0%
2 years, 6 months	2 SDs	82.87%	16.59%
	3 SDs	0%	35.34%
Formation Period, Trading period	***Opening trigger***	***Ubi_UniCredit***	***UniCredit_Ubi***
1 year, 6 months	2 SDs	19.95%	-12.24%
	3 SDs	25.90%	-6.51%
2 years, 6 months	2 SDs	22.48%	22.09%
	3 SDs	26.11%	20.94%
Formation Period, Trading period	***Opening trigger***	***Intesa_UniCredit***	***UniCredit_Intesa***
1 year, 6 months	2 SDs	19.01%	20.02%
	3 SDs	27.60%	29.82%
2 years, 6 months	2 SDs	27.73%	22.85%
	3 SDs	48.91%	42.01%

In this work for each pair under analysis both cointegration relationships with either one bank or the other as dependent variable are conducted. For each pair reported in the table, the first bank represents the time series that is used as dependent variable during the Engle-Granger test for cointegration, while the second bank represents the independent variable.

The fact that the prices at which we buy or sell the portfolio are more convenient considering the higher opening trigger can be the result of three different scenarios:

1. The price at which we purchase the undervalued stock is much lower at 3-SDs with respect to its price at 2-SDs, even if the price at which we sell the overvalued stock is slightly lower at 3-SDs compared to the its price at 2-SDs.
2. The price at which we sell the overvalued stock is much higher at 3-SDs with respect to its price at 2-SDs, even if the price at which we purchase the undervalued stock is slightly lower at 2-SDs compared to its price at 3-SDs.
3. The price at which we purchase the undervalued stock is lower at 3-SDs compared to its price at 2-SDs and the price at which we sell the overvalued stock is higher at 3-SDs compared to its price at 2-SDs.

Since the positions are closed in both cases either when the *z-score* reverts back to its equilibrium value, or on the last day of the trading period, having bought (or sold) the portfolio at a more convenient price allows us to obtain higher annualized returns.

However, one should notice that there exist five strategies for which an increase of the opening trigger produces a reduction of the annualized returns, and they are the following:

- *Intesa_BPM*, formation period of 1 year and trading period of 6 months.
- *BPM_Ubi*, formation period of 1 year and trading period of 6 months.
- *Ubi_Intesa*, formation period of 1 year and trading period of 6 months.
- *Intesa_Ubi*, formation period of 2 years and trading period of 6 months.
- *UniCredit_Ubi*, formation period of 2 years and trading period of 6 months.

As discussed in Section 3.4, the higher the opening trigger, the higher the deviation from the long-term equilibrium required to open a trade, and so the lower the number of openings and trades during a specific trading period. This is precisely the reason behind the reductions of the annualized returns for the aforementioned strategies: increasing the opening trigger generates a reduction of the number of profitable trades, which in turn produces a decline in the annualized returns (see *Table B1* in the *Appendix B*). *Figure 4.4* and *Figure 4.5* provide a graphical representation of the normalized spread for the five strategies during the 5-year time horizon considered. In particular, the panels on the left-hand side represent the normalized spreads with opening trigger set at 2-standard deviations, while the panels on the right-hand side represent the normalized spread with opening trigger set at 3-standard deviations. From these figures it is easy to observe that whenever the opening trigger increases, the number of openings and the number of profitable trades decrease, leading to a reduction of the returns. Moreover, it is also possible to see that when we increase the opening trigger from 2-SDs to 3-SDs, three out of five strategies, namely *BPM_Ubi*, *Intesa_Ubi* and *Ubi_Intesa*, never cross the opening trigger and so they never trade, leading to an annualized return of 0%.

Figure 4.4: *Normalized spread of Intesa_Ubi and UniCredit_Ubi with opening trigger set at 2-SDs*

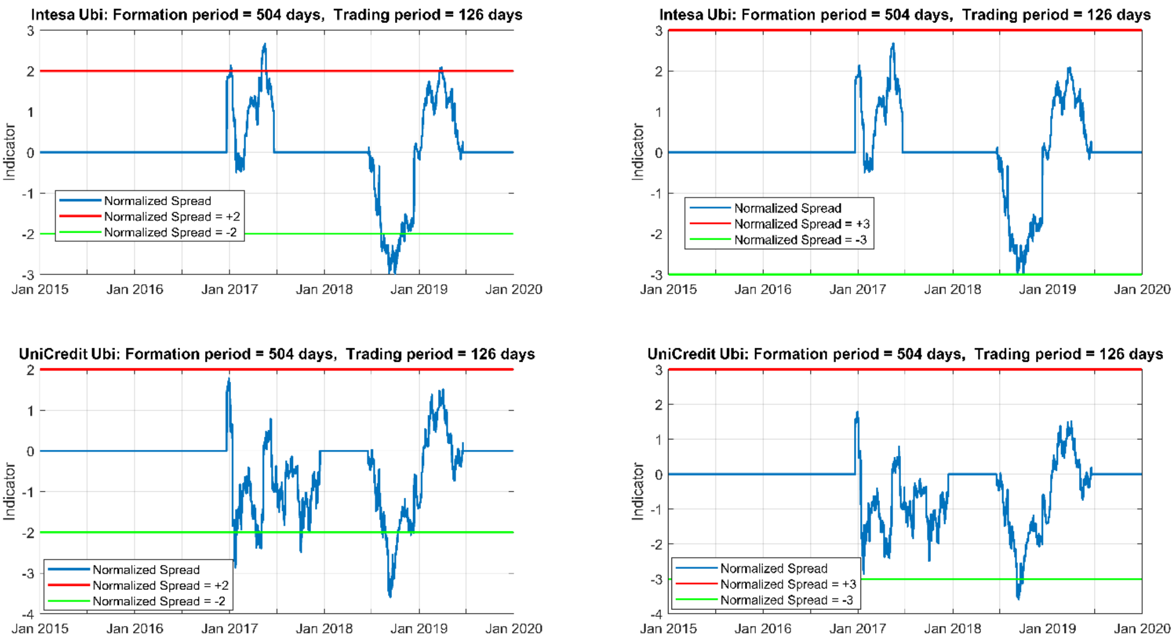
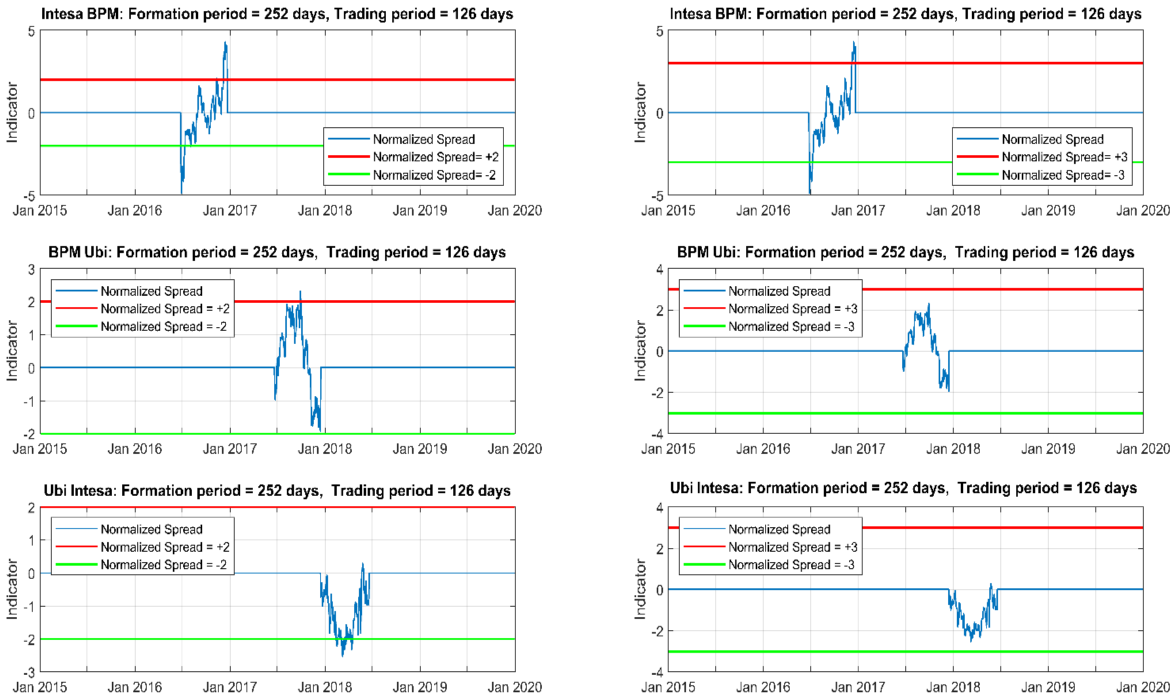


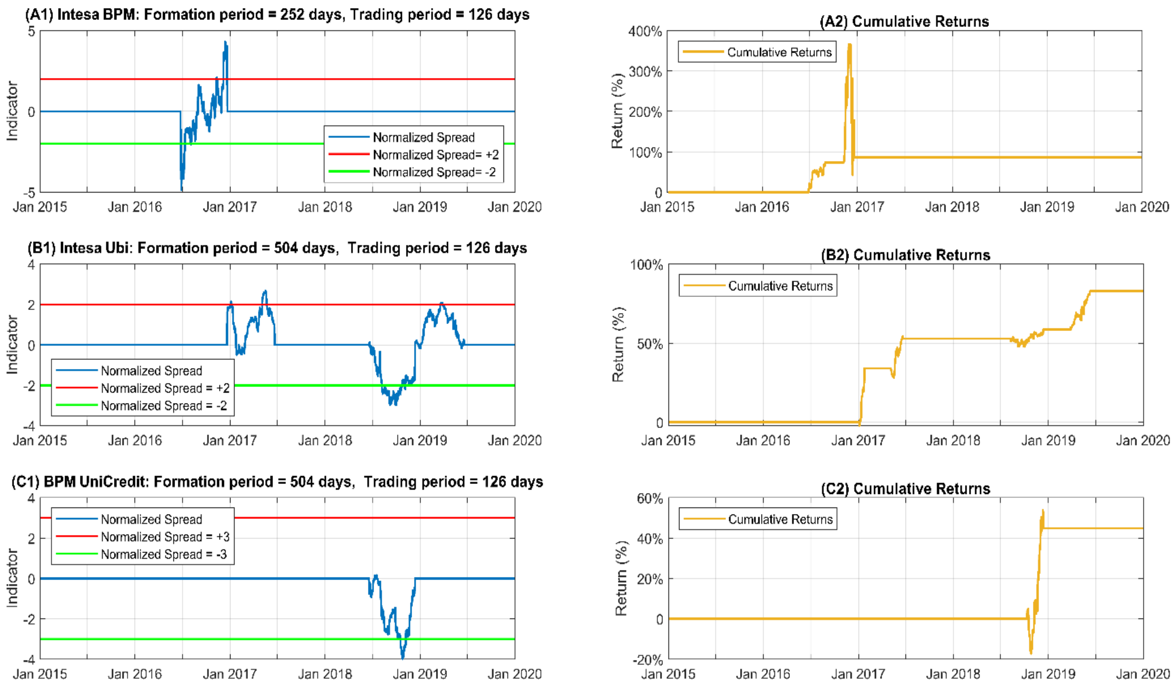
***Figure 4.5**: Normalized spread of Intesa_BPM, BPM_Ubi and Ubi_Intesa with opening trigger set at 2-SDs and 3-SDs.*



The most successful strategies are the following:

- *Intesa_BPM*, 1-year formation period and 6-month trading period with opening trigger set at 2-standard deviations, with an annualized return of 85.93% generated by 2 profitable trades and one unprofitable trade (see *Figure 4.6 A*).
- *Intesa_Ubi*, 2-year formation period and 6-month trading period with opening trigger set at 2-standard deviations, with an annualized return of 82.87%, generated by 4 profitable trades (see *Figure 4.6 B*).
- *BPM_UniCredit*, 2-year formation period and 6-month trading period with opening trigger set at 3-standard deviations, with an annualized return of 44.90%, produced by one profitable trade (see *Figure 4.6 C*).

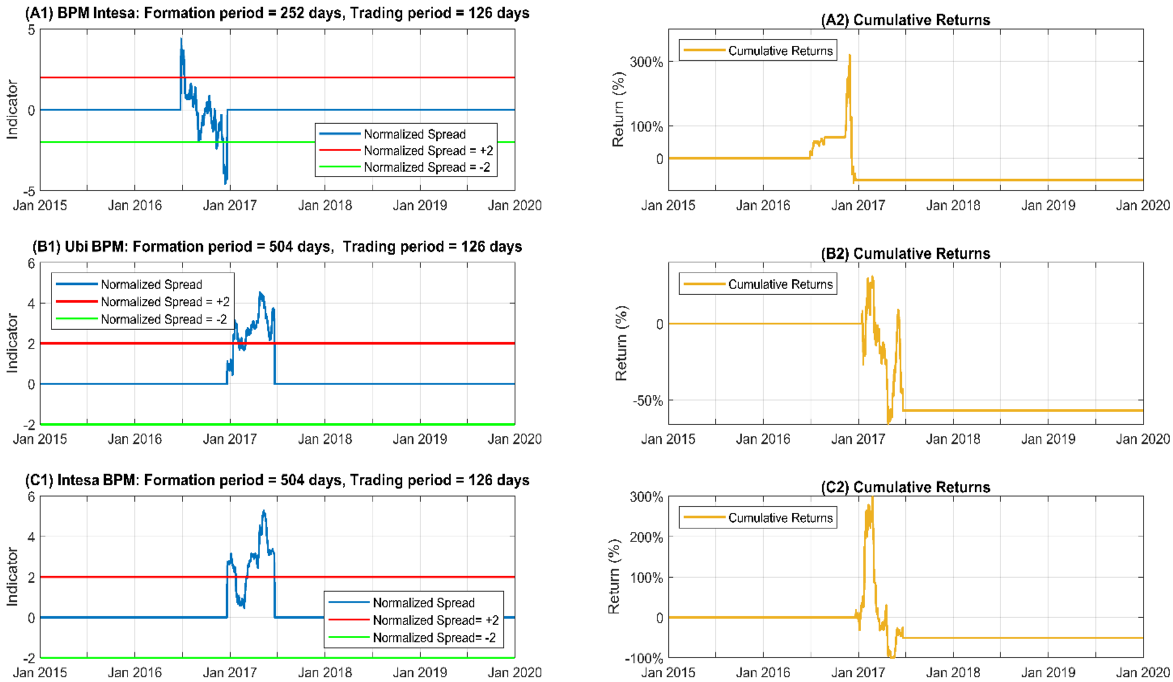
Figure 4.6: *Normalized spreads with related opening triggers and cumulative returns for the most successful strategies with closing trigger* $z_t = 0$



Instead, the least successful strategies are the following:

- *BPM_Intesa*, 1-year formation period and 6-month trading period with opening trigger set at 2-standard deviations, with an annualized return of -67.39% generated by one profitable trade and one unprofitable trade (see *Figure 4.7 A*).
- *Ubi_BPM*, 2-year formation period and 6-month trading period with opening trigger set at 2-standard deviations, with an annualized return of -56.98%, generated by one unprofitable trade (see *Figure 4.7 B*).
- *Intesa_BPM*, 2-year formation period and 6-month trading period with opening trigger set at 2-standard deviations, with an annualized return of -50.61%, produced by one unprofitable trade (see *Figure 4.7 C*).

__Figure 4.7__: Normalized spreads with related opening triggers and cumulative returns for the least successful strategies with closing trigger $z_t = 0$



In conclusion, in order to understand if the four parametrizations analyzed would have been profitable over the 5-year period considered, we decided to consider the average annualized return of the twelve pairs for each parametrization (i.e. formation period of 1 year and trading period of 6 months with opening trigger set at 2-SDs and 3SDs and formation period of 2 years and trading period of 6 months with opening trigger set at 2-SDs and 3-SDs), calculated as the sum of the annualized return of each strategy divided for the number of strategies which result to be cointegrated over the 5-year time horizon considered. The outcomes are summarized in *Table 4.8*. From these results we can conclude that all the parmetrizations considered result to be significantly profitable over the period from January 2015 and December 2019. In particular, the most succesful specification is 1-year formation period and 6-month trading period cominbed with a less restrictive opening trigger (2-SDs) with an average annualized return of 15.89%, follwed by the specification 2-year formation period and 6-month trading period, with opening trigger set at 3-SDs, which generates an average annualized return of 14.88%.

Table 4.8: *Average Annualized Returns with closing trigger:* $z_t = 0$.

Formation period, Trading period	***Opening trigger***	***Average Annualized Returns***	***Nr. of profitable str.***	***Nr. of unprofitable str.***	***Nr. of str. with zero returns***
1 year, 6 months	2 SDs	15.89%	6	2	0
	3 SDs	4.71%	3	3	2
2 years, 6 months	2 SDs	5.68%	8	4	0
	3 SDs	14.88%	7	4	1

Notice that the total number of strategies for the parametrization 1-year formation period and 6-month trading period is 8 instead of 12 because we are not considering the four pairs which result not to be cointegrated over the 5-year period considered.

4.4.2 Closing Trigger: $z_t > -q$ or $z_t < q$ with $q = 2, 3$

In this section we present the results obtained in the case in which the opened positions are closed when the normalized spread crosses the opening trigger twice (e.g. q-standard deviations), i.e. when $z_t > -q$ or $z_t < q$. The results reported in *Table 4.9* represent the annualized returns of the twelve pairs analysed, both considering a length of the formation period of one year (252 days) with trading period of 6 month (126 days), and a formation period of two years (504 days) with a trading period of 6 months (126 days). These parametrizations are analysed considering two opening triggers (and closing triggers) set at 2-standard deviations and 3-standard deviations, respectively.

As discussed in Section 4.3, if we consider the case in which a position is closed whenever the normalized spread crosses the opening trigger twice, the number of completed trades during a particular trading period will be greater (or equal) compared to the case in which the position is closed once the *z-score* reverts back to its long-term equilibrium. However, the potential profits for each completed trade will be lower. Since it is not possible to determine a priori which strategy is the most profitable, the purpose of this section is to present the results obtained when positions are closed upon re-convergence of the normalized spread to the opening trigger, i.e. the first time that the *z-score* crosses the opening trigger the position is opened and the second time the position is closed, and compare them with those presented in Section 4.4.1.

Table 4.9: Annualized returns with closing trigger: $z_t > -q$ or $z_t < q$ with $q = 2{,}3$

Formation Period, Trading period	***Opening trigger***	***BPM_Intesa***	***Intesa_BPM***
1 year, 6 months	2 SDs	-13.34%	-22.06%
	3 SDs	-42.17%	-33.32%
2 years, 6 months	2 SDs	-26.49%	-23.65%
	3 SDs	-3.64%	19.67%
Formation Period, Trading period	***Opening trigger***	***BPM_Ubi***	***Ubi_BPM***
1 year, 6 months	2 SDs	11.39%	Not cointegrated
	3 SDs	0%	Not cointegrated
2 years, 6 months	2 SDs	-22.03%	-57.06%
	3 SDs	11.84%	-15.97%
Formation Period, Trading period	***Opening trigger***	***BPM_UniCredit***	***UniCredit_BPM***
1 year, 6 months	2 SDs	Not cointegrated	Not cointegrated
	3 SDs	Not cointegrated	Not cointegrated
2 years, 6 months	2 SDs	14.74%	3.79%
	3 SDs	19.95%	30.17%
Formation Period, Trading period	***Opening trigger***	***Intesa_Ubi***	***Ubi_Intesa***
1 year, 6 months	2 SDs	Not cointegrated	15.81%
	3 SDs	Not cointegrated	0%
2 years, 6 months	2 SDs	11.32%	34.61%
	3 SDs	0%	2.45%
Formation Period, Trading period	***Opening trigger***	***Ubi_UniCredit***	***UniCredit_Ubi***
1 year, 6 months	2 SDs	26.28%	-7.17%
	3 SDs	4.83%	4.10%
2 years, 6 months	2 SDs	16.79%	17.17%
	3 SDs	15.83%	14.59%
Formation Period, Trading period	***Opening trigger***	***Intesa_UniCredit***	***UniCredit_Intesa***
1 year, 6 months	2 SDs	25.80%	34.26%
	3 SDs	42.91%	48.94%
2 years, 6 months	2 SDs	14.19%	17.72%
	3 SDs	10.02%	14.83%

In this work for each pair under analysis both cointegration relationships with either one bank or the other as dependent variable are conducted. For each pair reported in the table, the first bank represents the time series that is used as dependent variable during the Engle-Granger test for cointegration, while the second bank represents the independent variable.

Excluding the four pairs which result not to be cointegrated over the sample period considered, the annualized returns can be divided as follows:

- 65% positive annualized returns (26 strategies out of 40)
- 27.50% negative annualized returns (11 strategies out of 40)
- 7.50% zero annualized returns (3 strategies out of 40)

The positive returns range from 2.45% (*Ubi_Intesa*, formation period of 2 years and trading period of 6 months with opening trigger set at 3-standard deviations) to 48.94% (*UniCredit_Intesa*, formation period of 1 year and trading period of 6 months with opening trigger set at 3-standard deviations); the negative returns range from -3.64% (*BPM_Intesa*, formation period of 2 years and trading period of 6 months with opening trigger set at 3-standard deviations) to – 57.06% (*Ubi_BPM*, formation period of 2 years and trading period of 6 months with opening trigger set at 2-standard deviations).

The most successful strategies are the result of a very high number of profitable trades (see *Table B2* in the *Appendix B*) occurred during the period under analysis whose gains are not offset by the negative returns deriving from unprofitable trades (if any), and they are the following:

- *UniCredit_Intesa*, 1-year formation period and 6-month trading period with annualized returns of 48.94% (opening trigger set at 3-standard deviations), generated by 15 profitable trades (percentage of profitable trades: 100%), and 34.26% (opening trigger at 2-standard deviations), produced by 24 profitable trades (percentage of profitable trades: 96%). These normalized spread and cumulative returns of these strategies are represented in *Figure 4.8 A* and *Figure 4.8 B*, respectively.
- *Intesa_UniCredit*, 1-year formation period and 6-month trading period with an annualized return of 42.91% (opening trigger set at 3-standard deviations), generated by 17 profitable trades (percentage of profitable trades: 100%). The *z-score* and cumulative returns of this strategy are represented in *Figure 4.9 A*.
- *Ubi_Intesa*, 2-year formation period and 6-month trading period with an annualized return of 34.61% (opening trigger set at 2-standard deviations), produced by 12 profitable trades (percentage of profitable trades: 100%). A graphical representation of the normalized spread and the cumulative returns of this strategy can be found in *Figure 4.9 B*.

Figure 4.8: *Normalized spreads with related opening triggers and cumulative returns for the most successful strategies with closing trigger* $z_t > -q$ or $z_t < q$ with $q = 2,3$

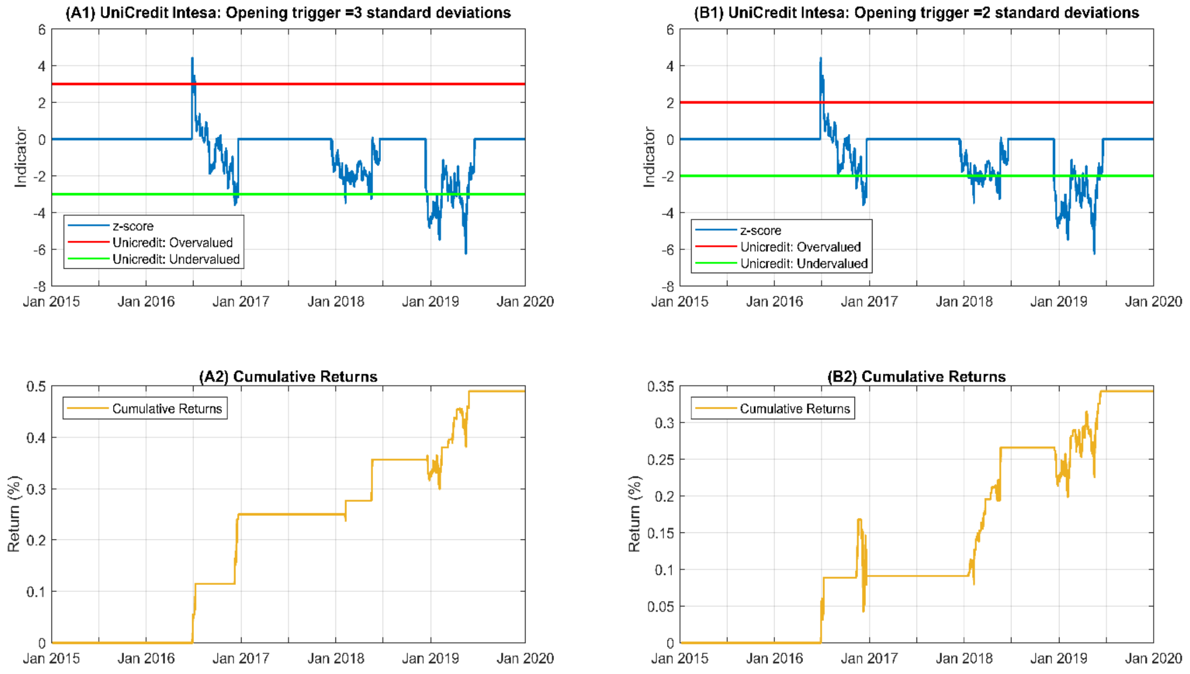
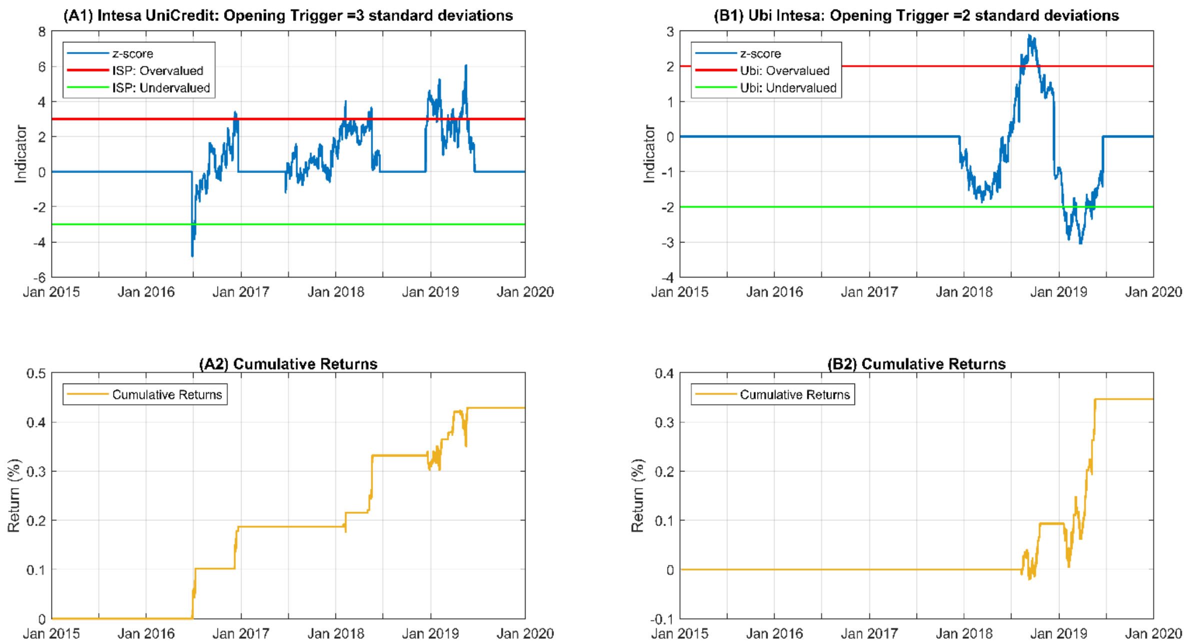


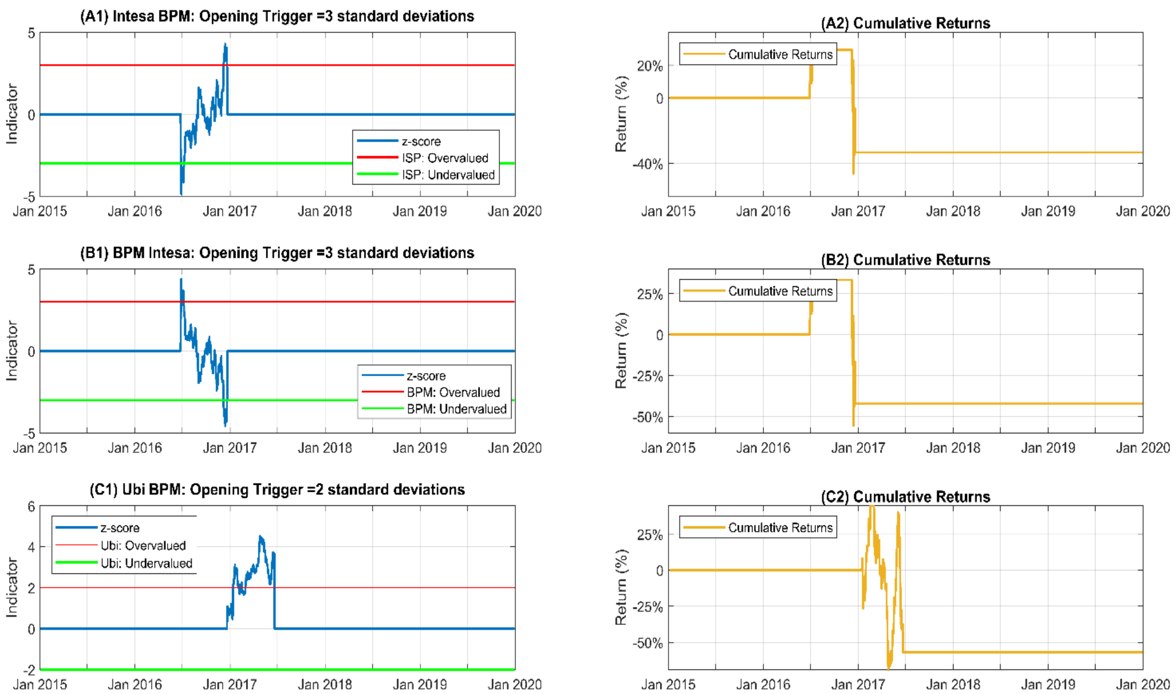
Figure 4.9: *Normalized spreads with related opening triggers and cumulative returns for the most successful strategies with closing trigger* $z_t > -q$ or $z_t < q$ with $q = 2,3$



Conversely, the negative annualized returns are the result of a very low number of profitable trades whose positive returns are completely offset by the negative returns deriving from unprofitable trades which occur on the last day of the trading period when all the opened positions are closed out (see *Table B2* in the *Appendix B*). The most unsuccessful strategies are the following:

- *Ubi_BPM*, 2-year formation period and 6-month trading period with an annualized return of -57.06% (with opening trigger set at 2-standard deviations), produced by 3 profitable trades and a single unprofitable trade (see *Figure 4.10 C*).
- *BPM_Intesa*, 1-year formation period and 6-month trading period with an annualized return of -42.17% (with opening trigger set at 3-standard deviations), resulting by 2 profitable trades and one unprofitable trade (see *Figure 4.10 B*).
- *Intesa_BPM*, 1-year formation period and 6-month trading period with an annualized return of -33.32% (with opening trigger set at 3-standard deviation), generated by one profitable trade and one unprofitable trade (see *Figure 4.10 A*).

Figure 4.10*: Normalized spread with related opening triggers and cumulative returns for the least successful strategies with closing trigger* $z_t > -q$ or $z_t < q$ with $q = 2{,}3$



Also in this case, in order to assess the profitability of the four parametrizations over the period considered, we consider the average annualized returns of the twelve pairs for each parametrization (i.e. formation period of 1 year and trading period of 6 months with opening trigger set at 2-SDs and 3SDs, formation period of 2 year and trading period of 6 months with opening trigger set at 2-SDs and 3-SDs). The results are summarized in *Table 4.10*. From these results it is possible to observe than only three parametrizations out of four are significantly profitable, while the specification 2-year formation period and 6-month trading period with opening trigger set at 2-standard deviations provides a slightly positive average annualized return of 0.09%, resulting from 4 unprofitable strategies (*BPM_Intesa, Intesa_BPM, BPM_Ubi* and *Ubi_BPM*) which almost completely offset the positive returns deriving from the 8 profitable strategies.

Table 4.10: *Average Annualized Returns with closing trigger:* $z_t > -q$ or $z_t < q$ with $q = 2{,}3$.

Formation period, Trading period	***Opening trigger***	***Average Annualized Returns***	***Nr. of profitable str.***	***Nr. of unprofitable str.***	***Nr. of str. with zero returns***
1 year, 6 months	2 SDs	8.87%	5	3	0
	3 SDs	3.16%	4	2	2
2 years, 6 months	2 SDs	0.09%	8	4	0
	3 SDs	9.98%	9	2	1

Notice that the total number of strategies for the parametrization 1-year formation period and 6-month trading period is 8 instead of 12 because we are not considering the four pairs which result not to be cointegrated over the 5-year period considered.

In this case the most succesful parametrization is 2-year formation period and 6-month trading period cominbed with an opening trigger of 3-SDs with an average annualized return of 9.98%. The reason behind the profitability of this parametrization is the high number of profitable strategies (9 out of 12) compared to the number of unprofitable strategies (2 out of 12) and to the number of strategies with zero returns (1 out of 12).

Comparing the results in *Table 4.10* with those summarized in *Table 4.8* it is possible to observe that pairs trading strategies are more profitable when we consider as closing trigger the convergence of the normalized spread to its long-term equilibrium, i.e. $z_t = 0$, since the average returns for all the parmetrizations analysed result to be greater.

4.4.3 Could a longer trading period reduce the impact of unprofitable trades?

The purpose of this section is to consider a single formation period of two years, from 2 January 2015 to 20 December 2016, and a single trading period of two years, from 21 December 2016 to 12 December 2018, to understand if the impact of the losses, due to the closure of the positions at the end of the 6-month trading period considered in Section 4.4.1 and in Section 4.4.2, could be reduced if a longer trading period is considered.

As discussed in Section 3.4, it does not exist a standard rule for deciding the lengths of the formation period and the trading period. However, the formation period should be long enough so that it is possible to verify whether a cointegration relationship between the two stocks under analysis exists or not, and the trading period should be chosen so that the selection process is recent, but not too long because it is possible that the cointegrating relationship estimated during the formation period may change during the trading period, making the pairs trading strategy completely unreliable. For these reasons, we must be aware that considering a relatively long trading period of two years could potentially lead to greater losses.

The results reported in *Table 4.11* represent the annualized returns of the twelve pairs for the case in which the pairs trading positions are closed upon convergence of the *z-score* to its long-term equilibrium, considering a length of the formation period of two year (504 days) with trading period of two years (504 days). As in the cases discussed above, the parametrization is analysed considering two opening triggers set at 2-standard deviations and 3-standard deviations, respectively.

Table 4.11: *Annualized returns for the parametrization 2-year formation period and 2-year trading period with closing trigger:* $z_t = 0$

Formation Period, Trading period	***Opening trigger***	***BPM_Intesa***	***Intesa_BPM***
2 years, 2 years	2 SDs	50.54%	50.77%
	3 SDs	50.91%	51.36%
Formation Period, Trading period	***Opening trigger***	***BPM_Ubi***	***Ubi_BPM***
2 years, 2 years	2 SDs	7.24%	7.96%
	3 SDs	13.13%	13.66%
Formation Period, Trading period	***Opening trigger***	***BPM_UniCredit***	***UniCredit_BPM***
2 years, 2 years	2 SDs	Not cointegrated	25.78%
	3 SDs	Not cointegrated	35.60%
Formation Period, Trading period	***Opening trigger***	***Intesa_Ubi***	***Ubi_Intesa***
2 years, 2 years	2 SDs	50.47%	Not cointegrated
	3 SDs	25.34%	Not cointegrated
Formation Period, Trading period	***Opening trigger***	***Ubi_UniCredit***	***UniCredit_Ubi***
2 years, 2 years	2 SDs	11.58%	10.46%
	3 SDs	5.12%	3.38%
Formation Period, Trading period	***Opening trigger***	***Intesa_UniCredit***	***UniCredit_Intesa***
2 years, 2 years	2 SDs	21.15%	19.96%
	3 SDs	25.50%	24.83%

Notice that all the annualized returns obtained with this particular specification result to be positive, with the only exceptions of *BPM_UniCredit* and *Ubi_Intesa* which result not to be cointegrated over the formation period and so they are not considered for the subsequent trading period. In particular, these positive returns range from 3.38% (*UniCredit_Ubi*, formation period of two years and trading period of two years with opening trigger set at 3-standard deviations) to 51.36% (*Intesa_BPM*, 2-year formation period and 2-year trading period with opening trigger at 3-standard deviations).

All the annualized returns stem from single profitable trades (see *Table B3* in the *Appendix B*), meaning that positions are opened only once during the entire trading period and then they are closed either because of the convergence of the normalized spread to the long-term equilibrium or on the last day of the trading period.

The only exceptions are represented by the following pairs:

- *Intesa_Ubi* (see *Figure 4.11 A1* and *Figure 4.11 A2*)
- *Ubi_UniCredit* (see *Figure 4.11 B1* and *Figure 4.11 B2*)
- *UniCredit_Ubi* (see *Figure 4.11 C1* and *Figure 4.11 C2*)

which under the less restrictive opening trigger (2-SDs) generate two profitable trades.

In *Figure 4.11* are represented the normalized spreads for the three aforementioned strategies, both considering an opening trigger of 2-standard deviations (panels on the left-hand side) and an opening trigger of 3-standard deviations (panels on the right-hand side). Moreover, as we expect, these are the only strategies whose annualized returns decrease when we increase the opening trigger from 2-standard deviations to 3-standard deviations.

In order to evaluate the profitability of the two parametrizations analysed in this section, i.e. 2-years formation period and 2-year trading period with opening triggers at 2-SDs and 3-SDs, we apply the same methodology used in Section 4.4.1 and 4.4.2. Thus, we compute the average annualized return of the ten pairs which result to be cointegrated over the formation period considered for each specification. The results are summarized in *Table 4.12.*

Figure 4.11: *Normalized spread with related opening for Intesa_Ubi, Ubi_UniCredit and UniCredit_Ubi*

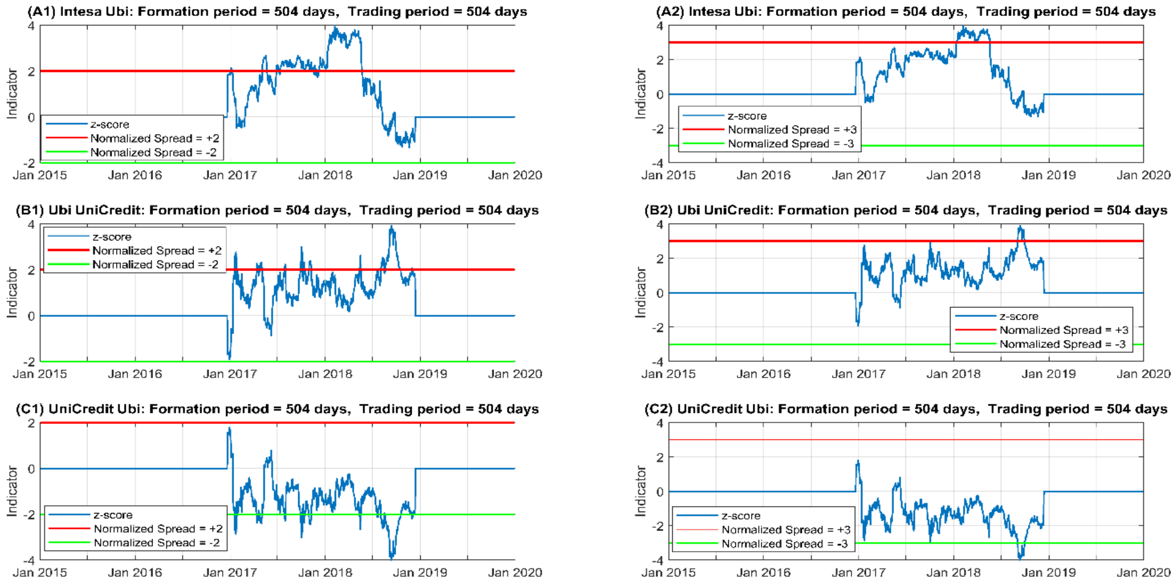


Table 4.12: *Average Annualized Returns for the parametrization 2-year formation period and 2-year trading period with closing trigger:* $z_t = 0$.

Formation period, Trading period	***Opening trigger***	***Average Annualized Returns***	***Nr. of profitable str.***	***Nr. of unprofitable str.***	***Nr. of str. with zero returns***
2 year, 2 years	2 SDs	25.59%	10	0	0
	3 SDs	24.88%	10	0	0

Notice that the total number of strategies for this parametrization is 10 instead of 12 because we are not considering the two pairs which result not to be cointegrated over the 2-year formation period considered, from 2 January 2015 to 20 December 2016.

From these results it is possible to observe that the two parametrizations provide very similar average annualized returns, which in both cases result to be economically significant. However, the specification with 2-standard deviations as opening trigger is slightly superior, with an average annualized return of 25.59%.

At this point, it could be useful to compare these results with those reported in *Table 4.8* to understand which specification among the six examined turns out to be the most profitable when we consider the convergence of the *z-score* to its long-term equilibrium as closing trigger. The average annualized returns obtained with a formation period and a trading period of two years, both considering 2-standard deviations and 3-standard deviations as opening trigger, clearly outperform those obtained in Section 4.4.1. The reason behind the success of these parametrization is the reduction of the number of unprofitable strategies, which in this particular case falls to zero.

Finally, we consider the case in which the pairs trading positions are closed when the normalized spread crosses the opening trigger twice, i.e. when $z_t > -q$ or $z_t < q$ with $q = 2{,}3$. Also in this case, the objective is to present the results obtained for the parametrization 2-year formation period and 2-year trading period and compare them with those summarized in *Table 4.10*, in order to understand if an increase of the trading period could improve the profitability of our strategy. The annualized returns for the twelve pairs under analysis in the case in which positions are closed upon re-convergence of the normalized spread to the opening trigger, considering a formation period of 2 years and a trading period of 2 years with opening trigger at 2-standard deviations and 3-standard deviations are reported in *Table 4.13*.

Table 4.13: *Annualized returns for the parametrization 2-year formation period and 2-year trading period with closing trigger*: $z_t > -q$ or $z_t < q$ with $q = 2{,}3$

Formation Period, Trading period	***Opening trigger***	***BPM_Intesa***	***Intesa_BPM***
2 years, 2 years	2 SDs	53.80%	85.28%
	3 SDs	79.43%	66.70%
Formation Period, Trading period	***Opening trigger***	***BPM_Ubi***	***Ubi_BPM***
2 years, 2 years	2 SDs	18.83%	23.73%
	3 SDs	24.69%	20.04%
Formation Period, Trading period	***Opening trigger***	***BPM_UniCredit***	***UniCredit_BPM***
2 years, 2 years	2 SDs	Not cointegrated	17.57%
	3 SDs	Not cointegrated	50.32%
Formation Period, Trading period	***Opening trigger***	***Intesa_Ubi***	***Ubi_Intesa***
2 years, 2 years	2 SDs	20.53%	Not cointegrated
	3 SDs	9.79%	Not cointegrated
Formation Period, Trading period	***Opening trigger***	***Ubi_UniCredit***	***UniCredit_Ubi***
2 years, 2 years	2 SDs	21.29%	20.94%
	3 SDs	0.79%	0.89%
Formation Period, Trading period	***Opening trigger***	***Intesa_UniCredit***	***UniCredit_Intesa***
2 years, 2 years	2 SDs	20.46%	21.80%
	3 SDs	13.17%	16.90%

Notice that, also in this case, all the annualized returns are positive and they range from 0.79% (*Ubi_UniCredit,* formation period of two years and trading period of two years with opening trigger set at 3-standard deviations) to 85.28% (*Intesa_BPM*, formation period of two years and trading period of two years with opening trigger set at 2-standard deviations). Looking at the results stated in *Table B4* in the *Appendix B*, it is possible to observe that these positive annualized returns are the result of multiple profitable trades and no unprofitable trades occurred during the trading period examined. Furthermore, the most successful strategies are associated with a relatively high number of profitable trades, while the least successful are linked to a low number of profitable trades. In particular, the strategies with the highest returns are the following:

- *Intesa_BPM* with annualized returns of 85.28% (generated by 4 profitable trades) under the less selective opening scheme (2-SDs), and 66.70% (produced by 9 profitable trades) under the 3-SDs rule.
- *BPM_Intesa* with an annualized return of 79.43% (generated by 10 profitable trades) under the 3-SDs opening rule.

Instead, the least successful strategies can be summarized as follows:

- *Ubi_UniCredit* with an annualized return of 0.79% (produced by 1 profitable trade) under the 3-SDs rule.
- *UniCredit_Ubi* with an annualized return of 0.89% (produced by 2 profitable trades) under the more selective opening scheme (3-SDs).
- *Intesa_Ubi* with an annualized return of 9.79% (generated by 2 profitable trades) under the 3-SDs opening rule.

In the end, we proceed in calculating the overall performance of these parametrizations by considering the average annualized returns of the ten pairs which result to be cointegrated over the formation period considered for each specification. The results are reported in *Table 4.14.*

Table 4.14: *Average Annualized Returns for the parametrization 2-year formation period and 2-year trading period with closing trigger:*$z_t > -q$ or $z_t < q$ with $q = 2,3$.

Formation period, Trading period	***Opening trigger***	***Average Annualized Returns***	***Nr. of profitable str.***	***Nr. of unprofitable str.***	***Nr. of str. with zero returns***
2 year, 2 years	2 SDs	30.42%	10	0	0
	3 SDs	28.27%	10	0	0

Notice that the total number of strategies for this parametrization is 10 instead of 12 because we are not considering the two pairs which result not to be cointegrated over the 2-year formation period considered, from 2 January 2015 to 20 December 2016.

Notice that both the parametrizations provide economically significant annualized returns, which are the result of ten profitable strategies without any unprofitable strategies or strategies yielding zero returns. Specifically, the best parametrization is the one in which we consider a less selective opening scheme (2-SDs) which generates a remarkable average annualized return of 30.42%.

Comparing these results with those reported in *Table 4.10* it is evident that the returns achieved considering a longer trading period strongly outperform those realized with a trading period of 6 months, either by examining a formation period of one year or two years. Also in this case, the greater results obtained considering a longer trading period are due to the fact that with this particular combination of formation period (2 years) and trading period (2 years) we completely eliminate the unprofitable strategies. However, it is important to highlight that the aim of this work is of course not to find the 'optimal' parametrization of pairs trading strategies. The objective was just to establish if considering a sufficiently long trading horizon the impact of the losses due to the closure of the positions at the end of a relatively short trading period of 6 months could be reduced, and it turns out that this is exactly what happens in our case, since considering a trading period of two years we are able to completely eliminate unprofitable trades and so achieving greater returns for both the closing schemes examined.

Conclusions

The purpose of this thesis was to investigate the profitability of different cointegration-based pairs trading strategies, using the daily closing stock prices of the major banks in the Italian banking system over the period from 2 January 2015 to 30 December 2019.

First of all, we provide an accurate overview of the various classes of statistical arbitrage strategies, with particular emphasis on the different pairs trading approaches cited in the literature. Then we proceed with presenting all the statistical and econometric elements required to understand the concept of cointegration, as exposed by Engle and Granger in 1987. Finally, we provide a rigorous analysis of the three-step process (i.e. pre-selection of the stock pairs, testing for cointegration and trading design) that need to be followed to implement a cointegration-based pairs trading strategy.

For the purpose of this thesis we have followed the work by Huck and Afawubo (2015), which consider two different lengths of the formation period (1 year and 2 years) and a single length of the trading period (6 months) combined in two different strategies which are 1-year formation period and 6-month trading period and 2-year formation period and 6-month trading period, respectively. Our findings indicate that almost all these strategies are significantly profitable either examining a more selective opening scheme (3-standard deviations) and a less selective one (2-standard deviations), both considering as closing trigger the convergence of the normalized spread to its long-term equilibrium and the re-convergence of the spread to the opening trigger (i.e. the first time that the spread crosses the opening trigger the position is opened and the second time the position is unwound). The only exception is represented by the parametrization 2-year formation period and 6-month trading period with opening trigger set at 2-standard deviations considering as closing trigger the re-convergence of the normalized spread to the opening trigger which, over the 5-year trading horizon considered, generates a slightly positive average annualized return of 0.09% (see *Table 4.10*). Moreover, we show that pairs trading strategies are more profitable when we consider as closing trigger the convergence of the normalized spread to its long-term equilibrium, i.e. $z_t = 0$, since the average returns for all the parmetrizations analysed result to be greater. In particular, for these kind of specification the average annualized returns range from 4.71% considering a formation period of 1 year and a trading period of 6 months with an opening trigger of 3-standard deviations to 15.89% considering a formation period of 1 year and a trading period of 6 months with an opening trigger of 2-standard deviations (see *Table 4.8*).

Furthermore, we have decided to investigate if the impact of losses, due to the closure of the pairs trading positions at the end of the 6-month trading period before the convergence to the closing trigger has occurred, could be reduced if a longer trading period was taken into account. To this effect, we have considered a single formation period of two years, from 2 January 2015 to 20 December 2016, and a single trading period of two years, from 21 December 2016 to 12 December 2018. From this analysis it turns out that considering a longer trading period generates greater average annualized returns with respect to those obtained with a shorter trading period of 6 months. The reason behind the success of these specifications is the reduction of the number of unprofitable strategies, which in these particular cases falls to zero. The main limitation of this work is that we focus only on the analysis of annualized returns and we do not consider any measure of risk, which would have provided a more complete description of the profitability of the strategies analyzed. For example, one of the main metrics used in the literature to assess the riskiness of a pairs trading strategy is the maximum drawdown. According to Caldeira and Moura (2013), 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. The maximum drawdown (*MDD*) defines the total percentage loss experienced by a pairs trading strategy before it starts winning again. In other words, it is the maximum negative distance between a local maximum and the subsequent local minimum, and it gives a good measure of the downside risk for the investor (Dunis *et al*., 2010). Analytically it can be defined as (Caldeira and Moura, 2013):

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

where R_t is the daily return of the portfolio of pairs under analysis on day *t*.

Another way of improving this research could be to consider some risk management tools as an attempt to reduce the risk and limit the downside of the pairs trading strategy. In our work we have tried to improve the performance of the strategy by optimizing the pairs selection process and the length of the trading period. However, it is possible to improve the strategy by managing in a more efficient way the risk of the portfolio. In Section 3.4, we have highlighted the fact that one of the main risks associated to pairs trading strategies is that securities' prices may continue to diverge after position opening rather than revert to the long-term equilibrium, potentially leading to substantial losses for the trader. Thus, in order to limit the decline in the profitability of the strategy it is possible to implement proper risk management tools, such as:

- *Stop loss constraint*: which is a function used to automatically unwind a position whenever a pre-defined loss is registered.

- *Maximum holding length of a trade*: which is the maximum time a trade can be kept opened, exceeding this threshold automatically generates an exit signal.

Finally, in order to achieve a more realistic and rigorous modelling of pairs trading strategies, it could be useful to consider models in which the cointegration coefficient is not constant but may vary with time. This can be done using the Kalman filter, which allows parameters (in this case the cointegration coefficient) to vary over time (Dunis *et al.*, 2010). According to Hamilton (1994, p. 372), the idea behind the concept of Kalman filter is to express a dynamic system in a particular form called *state-space representation.* The Kalman filter is an algorithm for sequentially updating a linear projection for the system, providing a way to calculate exact finite-sample forecasts and the exact likelihood function for Gaussian ARMA process. The objective of state space modelling is to provide an estimate of the unobservable states of the dynamic system in the presence of noise. The Kalman filter is a recursive method for filtering out the observations noise in order to optimally estimate the state space vector at a generic time *t*, based on the information available at time *t*. In other words, the filter consists of a system of equations which allow us to update the estimate of a state when new observations become available (Bentz, 2003). In their seminal paper Elliott *et al.* (2005) suggest that Gaussian linear state-space processes may be suitable for modelling spreads arising in pairs trading, and they describe how such models can yield statistical arbitrage strategy. As argued by Miao (2014), the Kalman filter can be used, within the pairs trading framework, for estimating and identifying the adaptive stability of the parameters of the cointegration model in real-time mode, providing enhancements of profitability and mitigation of risks for pairs trading based on the market-neutral statistical arbitrage strategy.

References

R. Ahmad, P. Willmott (2005), *Which Free Lunch Would You Like Today Sir? Delta-Hedging, Volatility Arbitrage and Optimal Portfolios*, Willmott Magazine, 2005: 64-79.

C. Alexander, I. Giblin, W. Weddington III (2001), *Cointegration and asset allocation: a new active hedge fund strategy*, Discussion Papers in Finance ISMA Centre 2001-03, University of Reading.

C. Alexander, A. Dimitriu (2005), *Indexing and Statistical Arbitrage*, The Journal of Portfolio Management 31(2): 50-63.

C. Alexander (2008), *Practical Financial Econometrics*, John Wiley & Sons, Chichester (UK).

M. Avellaneda, J. H. Lee (2010), *Statistical Arbitrage in the US equities market*, Quantitative Finance 10: 1-22.

K. Belkhayat (2017), *Statistical arbitrage: systematic equity pairs trading*, thesis, ETH Zurich.

C. Brooks (2008), *Introductory Econometrics for Finance*, 2nd edition, Cambridge University Press.

Y. Bentz (2003), *Quantitative Equity Investment Management with Time-Varying Factor Sensitivities*, in C. Dunis, J. Laws, P. Naim, *Applied Quantitative Methods for Trading and Investment*, John Wiley & Son, Chichester: 213-237.

J. P. Broussard, M. Vaihekoski (2012), *Profitability of Pairs Trading in an Illiquid Market with Multiple Share Classes*, Journal of International Financial Markets, Institutions and Money 22(5): 1188-1201

J. F. Caldeira, G. V. Moura (2013), *Selection of a Portfolio of pairs based on Cointegration: a Statistical Arbitrage Strategy*, Brazilian Review of Finance 11(1): 49-80.

E. P. Chan (2013), *Algorithmic Trading: Winning Strategies and Their Rationale*, John Wiley & Sons, Hoboken (NJ).

H. J. Chen, S. J. Chen, Z. Chen, F. Li (2019), *Empirical Investigation of an Equity Pairs Trading Strategy*, Management Science 65(1): 370-389.

R. Davidson, J. G. MacKinnon (2004), *Econometric Theory and Methods*, Oxford University Press (NY).

D. A. Dickey, W. A. Fuller (1979), *Distribution of the Estimators for Autoregressive Time Series with a Unit Root*, Journal of the American Statistical Association, 74: 427-431.

B. Do, R. Faff, K. Hamza (2006), *A new approach to modelling and estimation for pairs trading*, Proceedings of 2006 Financial Management Association European Conference.

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

B. Do, R. Faff (2012), *Are Pairs Trading Profits Robust to Trading Costs?*, Journal of Financial Research 35:261-287.

J. Duarte, F.A. Longstaff, F. Yu (2006), *Risk and Return in Fixed Income Arbitrage: Nicklers in Front of a Steamroller?* The Review of Financial Studies, 20: 769-811.

C. L. Dunis, R. Ho (2005), *Cointegration Portfolios of European Equities for Index Tracking and Market Neutral Strategies*, Journal of Asset Management 6(1): 33-52.

R. J. Dunis, C. Giorgioni, J. Laws (2010), *Statistical Arbitrage and High-Frequency Data with an Application to Eurostoxx 50 equities*, Working paper, Liverpool Business School.

R.J. Elliott, J. Van Der Hoek, W.P. Malcolm (2005), *Pairs trading*, Quantitative Finance 5(3): 271-276.

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

A. Fiz (2014), *Pairs Trading: An Empirical Study*, Master Thesis, University of the Basque Country, Complutense University of Madid, University of Castilla La Mancha and University of Valencia.

W. Fuller (1976), *Introduction to Statistical Times Series*, Wiley, New York.

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

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

C. W. J. Granger (1981), *Some Properties of Time Series Data and Their Use in Econometric Model Specification*, Journal of Econometrics, 121-130.

C. W. J. Granger, P. Newbold (1974), *Spurious Regressions in Econometrics*, Journal of Econometrics, 2: 111-120.

C. W. J. Granger, A. A. Weiss (1983), *Time Series Analysis of Error-Correcting Models*, Studies in Econometrics, Time Series, and Multivariate Statistics, New York: Academic Press, 255-278.

W.H. Greene (2018), *Econometric Analysis*, 8th edition, Pearson Education Inc (NY).

J. Hamilton (1994), *Time series analysis*, Princeton University Press.

N. Huck, K. Afawubo (2015), *Pairs Trading and Selection Methods: Is Cointegration Superior?*, Applied Economics 47(6): 599-613.

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

S. Johansen (1991), *Estimation and Hypothesis Testing of Cointegration Vectors in Gaussian Vector Autoregressive Models*, Econometrica, 59: 1551-1580.

J.W. Jurek, H. Yang (2007), *Dynamic portfolio selection in arbitrage*, Working paper, Harvard University.

G. Kirchgässner, J. Wolters, U. Hassler (2013), *Introduction to Modern Time Series Analysis,* 2nd edition, Springer (Berlin).

E. Kočenda, A. Černý (2015), *Elements of Time Series Econometrics: An Applied Approach*, Karolinum Press (Prague).

C. Krauss, J. Stübinger (2015), *Nonlinear dependence modelling with bivariate copulas: statistical arbitrage pairs trading on the S&P 100*, IWQW Discussion Paper Series, University of Erlangen-Nürnberg, Nürnberg.

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

D. Kwiatkowski, P. C. B. Phillips, P. Schmidt, Y. Shin (1992), *Testing the Null Hypothesis of Stationarity Against the Alternative of a Unit Root*, Journal of Econometrics, 54: 159-178.

M. Lazzarino, J. Berrill, A. Šević (2018), *What is Statistical Arbitrage?,* Theoretical Economics Letters, 8: 888-908.

R.Q. Liew, Y. Wu (2013), *Pairs trading: a copula approach,* Journal of Derivatives & Hedge Funds 19(1): 12-30.

Y.X. Lin, M. McCrae and C. Gulati (2006), *Loss protection in pairs trading through minimum profit bounds: a cointegration approach*, Journal of Applied Mathematics and decision Sciences 2006: 1-14.

J. Liu, A. Timmermann (2013), *Optimal convergence trade strategies*, Review of Financial Studies 26(4):1048-1086.

J. G. MacKinnon (1991), *Critical Values for Cointegration Tests,* in: R. F. Engle, C. W. J. Granger (eds.), *Long-Run Economic Relationship: Readings in Cointegration,* Oxford University Press (NY) pp. 266-276.

H. M. Markowitz (1952), *Portfolio Selection*, The Journal of Finance 7(1): 77-91.

Mediobanca (2019), *Focus sul Sistema Bancario Italiano (dati 2018*), Area Studi Mediobanca.

G. J. Miao (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.

M. Mori, A.J. Ziobrowski (2011), *Performance of pairs trading strategy in the U.S. REIT market.* Real Estate Economics 39(3): 409-428.

A. Naccarato, A. Pierini, G. Ferraro (2019), *Markowitz Portfolio Optimization through Pairs Trading Cointegrated Strategy in Long-Term Investment*, Annals of Operations Research.

R.B. Nelesen (2006), *An introduction to copulas,* 2nd edition, Springer Science Business Media, Inc.

M. Osterwald-Lenum (1992), *A Note with Quantiles of the Asymptotic Distribution of the Maximum Likelihood Cointegration Rank Test Statistics*, Oxford Bulletin of Economics and Statistics 54: 461-472.

K. Patterson (2010), *A Primer for Unit Root Testing*, Basingstoke: Palgrave Macmillan.

P. C. B. Phillips (1986), *Understanding Spurious Regressions in Econometrics*, Journal of Econometrics, 33: 311-340.

P. C. B. Phillips (1987), *Time Series Regression with a Unit Root*, Econometrica 55: 277-301.

P. C. B. Phillips, P. Perron (1988), *Testing for a Unit Root in Time Series Regression*, Biometrika, 75: 335-345.

A. Pole (2008), *Statistical arbitrage: algorithmic trading insights and techniques*, John Wiley & Sons, Hoboken (NJ).

H. Puspaningrum (2012), *Pairs trading using cointegration approach.* PhD thesis, University of Wollongong.

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

S. O. Rice (1944), *Mathematical Analysis of Random Noise*, Bell System Technical Journal 23: 282-332.

S.A. Ross (2004), *Neoclassical Finance*, Princeton University Press, Princeton (NJ).

S. E. Said, D. A. Dickey (1984), *Testing for Unit Roots in Autoregressive Moving Average Models of Unknown Order*, Biometrika, 71: 599-607.

P. Saks, D. Maringer (2008), *Genetic programming in statistical arbitrage,* Proceedings of the 2008 conference on Applications of evolutionary computing, 73-82.

S. Schaefer, I. Strebulaev (2004), *Capital Structure Arbitrage: An Empirical Investigation of Risk, Return and Hedging,*

W.F. Sharpe, G.J. Alexander, J.V. Bailey (1999), *Investments,* 6th edition, Prentice Hall, Upper Saddle River (NJ).

A. Shleifer, R.W. Vishny (1997), *The Limits of Arbitrage*, The Journal of Finance, 52:35-55.

M. Sklar (1959), *Fonctions de réparticion à n dimensions et leurs marges*, Publications de l'Institut de Statistique de L'Universitè de Paris (8): 229-231.

R. Ssekuma (2011), *A Study of Cointegration Models with Applications*, Master Thesis in Statistics, University of South Africa, South Africa.

Y. Stander, D. Marais, I. Botha (2013), *Trading strategies with copulas,* Journal of Economic and Financial Sciences 6(1): 83-107.

F. Stefanini (2006), *Investment Strategies of Hedge Funds*, John Wiley & Sons, Hoboken (NJ).

J. H. Stock, M. W. Watson (1988), *Testing for Common Trends*, Journal of the American Statistical Association, 83: 1097-1107.

K. Triantafyllopoulos, G. Montana (2011), *Dynamic modelling of mean-reverting spreads for statistical arbitrage*, Computational Management Science 8(1-2): 23-49.

M. Verbeek (2017), *A Guide to Modern Econometrics*, 5th edition, John Wiley & Sons, Rotterdam.

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

H. White (2000), *A reality check for data snooping*, Econometrica 68(5): 1097-1126.

W. Xie, Q.R. Liew, Y. Wu, X. Zou (2014), *Pairs trading with copulas,* Working paper, Nanyang Technological University.

E. Zivot, J. Wang (2003), *Modelling Financial Time Series with S-PLUS*, Springer (NY).

Appendices

Appendix A: Statistical Tables

Table A1: Critical Values for the Phillips-Perron $\boldsymbol{Z_\pi}$ Tests and for the Dickey-Fuller Test Based on Estimated OLS Autoregressive Coefficient

Sample Size T	Probability that $T(\hat{\phi}_T - 1)$ is less than entry							
	0.01	0.025	0.05	0.10	0.9	0.95	0.975	0.99
			Case 1					
25	-11.90	-9.30	-7.30	-5.30	1.01	1.40	1.79	2.28
50	-12.90	-9.90	-7.70	-5.50	0.97	1.35	1.70	2.16
100	-13.30	-10.20	-7.90	-5.60	0.95	1.31	1.65	2.09
250	-13.60	-10.30	-8.00	-5.70	0.93	1.28	1.62	2.04
500	-13.70	-10.40	-8.00	-5.70	0.93	1.28	1.61	2.04
∞	-13.80	-10.50	-8.10	-5.70	0.93	1.28	1.60	2.03
			Case 2					
25	-17.20	-14.60	-12.50	-10.20	-0.76	0.01	0.65	1.40
50	-18.90	-15.70	-13.30	-10.70	-0.81	-0.07	0.53	1.22
100	-19.80	-16.30	-13.70	-11.00	-0.83	-0.10	0.47	1.14
250	-20.30	-16.60	-14.00	-11.20	-0.84	-0.12	0.43	1.09
500	-20.50	-16.80	-14.00	-11.20	-0.84	-0.13	0.42	1.06
∞	-20.70	-16.90	-14.10	-11.30	-0.85	-0.13	0.41	1.04
			Case 4					
25	-22.50	-19.90	-17.90	-15.60	-3.66	-2.51	-1.53	-0.43
50	-25.70	-22.40	-19.80	-16.80	-3.71	-2.60	-1.66	-0.65
100	-27.40	-23.60	-20.70	-17.50	-3.74	-2.62	-1.73	-0.75
250	-28.40	-24.40	-21.30	-18.00	-3.75	-2.64	-1.78	-0.82
500	-28.90	24.80	-21.50	-18.10	-3.76	-2.65	-1.78	-0.84
∞	-29.50	-25.10	-21.80	-18.30	-3.77	-2.66	-1.79	-0.87

The probability shown at the head of the column is the area in the left-hand tail.

Source: Wayne A. Fuller (1976, p. 371), *Introduction to Statistical Time Series,* Wiley, New York.

Table A2: Critical Values for the Phillips-Perron $\boldsymbol{Z_\tau}$ Tests and for the Dickey-Fuller Test Based on Estimated OLS *t* Statistic

Sample Size T	Probability that $(\hat{\phi}_T - 1)/\hat{\sigma}_{\hat{\phi}_T}$ is less than entry							
	0.01	0.025	0.05	0.10	0.9	0.95	0.975	0.99
				Case 1				
25	-2.66	-2.26	-1.95	-1.60	0.95	1.33	1.70	2.16
50	-2.62	-2.25	-1.95	-1.61	0.91	1.31	1.66	2.08
100	-2.60	-2.24	-1.95	-1.61	0.90	1.29	1.64	2.03
250	-2.58	-2.23	-1.95	-1.62	0.89	1.29	1.63	2.01
500	-2.58	-2.23	-1.95	-1.62	0.89	1.28	1.62	2.00
∞	-2.58	-2.23	-1.95	-1.62	0.89	1.28	1.62	2.00
				Case 2				
25	-3.75	-3.33	-3.00	-2.63	-0.37	0.00	0.34	0.72
50	-3.58	-3.22	-2.93	-2.60	-0.40	-0.03	0.29	0.66
100	-3.51	-3.17	-2.89	-2.58	-0.42	-0.05	0.26	0.63
250	-3.46	-3.14	-2.88	-2.57	-0.42	-0.06	0.24	0.62
500	-3.44	-3.13	-2.87	-2.57	-0.43	-0.07	0.24	0.61
∞	-3.43	-3.12	-2.86	-2.57	-0.44	-0.07	0.23	0.60
				Case 4				
25	-4.38	-3.95	-3.60	-3.24	-1.14	-0.80	-0.50	-0.15
50	-4.15	-3.80	-3.80	-3.18	-1.19	-0.87	-0.58	-0.24
100	-4.04	-3.73	-3.73	-3.15	-1.22	-0.90	-0.62	-0.28
250	-3.99	-3.69	-3.69	-3.13	-1.23	-0.92	-0.64	-0.31
500	-3.98	-3.68	-3.68	-3.13	-1.24	-0.93	-0.65	-0.32
∞	-3.96	-3.66	-3.66	-3.12	-1.25	-0.94	-0.66	-0.33

The probability shown at the head of the column is the area in the left-hand tail.

Source: Wayne A. Fuller (1976, p. 371), *Introduction to Statistical Time Series*, Wiley, New York

Table A3: KPSS Critical Values

Critical Level	Critical Value	
	For the Stationary Case ($D_t = 1$)	*For the Trend-Stationary case ($D_t = (1, t)'$)*
0.10	0.347	0.119
0.05	0.463	0.146
0.025	0.574	0.176
0.01	0.739	0.216

Source: D. Kwiatkowski, P.C.B. Phillips, P. Schmidt, Y. Shin (1992).

Table A4: Critical Values for Johansen's Likelihood Ratio Test of the Null Hypothesis of h Cointegrating Relations Against the Alternative of No Restrictions

Number of random walks ($g = n - h$)	Sample size (T)	Probability that $2(\mathcal{L}_A - \mathcal{L}_0)$ is greater than entry					
		0.500	0.200	0.100	0.050	0.025	0.010
		Case 1					
1	400	0.58	1.82	2.86	3.84	4.93	6.51
2	400	5.42	8.45	10.47	12.53	14.43	16.31
3	400	14.30	18.83	21.63	24.31	26.64	29.75
4	400	27.10	33.16	36.58	39.89	42.30	45.58
5	400	43.79	51.13	55.44	59.46	62.91	66.52
		Case 2					
1	400	2.415	4.905	6.691	8.083	9.658	11.576
2	400	9.335	13.038	15.583	17.844	19.611	21.962
3	400	20.188	25.445	28.436	31.256	34.062	37.291
4	400	34.873	41.623	45.248	48.419	51.801	55.551
5	400	53.373	61.566	65.956	69.977	73.031	77.911
		Case 4					
1	400	0.447	1.699	2.816	3.962	5.332	6.936
2	400	7.638	11.164	13.338	15.197	17.299	19.310
3	400	18.759	23.868	26.791	29.509	32.313	35.397
4	400	33.672	40.250	43.964	47.181	50.424	53.792
5	400	52.588	60.215	65.063	68.905	72.140	76.955

The probability shown at the head of the column is the area in the right-hand tail. The number of random walks under the null hypothesis (g) is given by the number of variables described by the vector autoregression (n) minus the number of cointegrating relations under the null hypothesis (h). In each case the alternative is $g = 0$.

Source: Michael Osterwald-Lenum (1992), A Note with Quantiles of the Asymptotic Distribution of the Maximum Likelihood Cointegration Rank Test Statistics, Oxford Bulletin of Economics and Statistics 54: p. 462.

Table A5: Critical Values for Johansen's Likelihood Ratio Test of the Null Hypothesis of h Cointegrating Relations Against the Alternative of $h+1$ Relations

Number of random walks ($g = n - h$)	Sample size (T)	Probability that $2(\mathcal{L}_A - \mathcal{L}_0)$ is greater than entry					
		0.500	0.200	0.100	0.050	0.025	0.010
	Case 1						
1	400	0.58	1.82	2.86	3.84	4.93	6.51
2	400	4.83	7.58	9.52	11.44	13.27	15.69
3	400	9.71	13.31	15.59	17.89	20.02	22.99
4	400	14.94	18.97	21.58	23.80	26.14	28.82
5	400	20.16	24.83	27.62	30.04	32.51	35.17
	Case 2						
1	400	2.415	4.905	6.691	8.083	9.658	11.576
2	400	7.474	10.666	12.783	14.595	16.403	18.782
3	400	12.707	16.521	18.959	21.279	23.362	26.154
4	400	17.875	22.341	24.917	27.341	29.599	32.616
5	400	23.132	27.953	30.818	33.262	35.700	38.858
	Case 4						
1	400	0.447	1.699	2.816	3.962	5.332	6.936
2	400	6.852	10.125	12.099	14.036	15.810	17.936
3	400	12.381	16.324	18.697	20.778	23.002	25.521
4	400	17.719	22.113	24.712	27.169	29.335	31.943
5	400	23.211	27.899	30.774	33.178	35.546	38.341

The probability shown at the head of the column is the area in the right-hand tail. The number of random walks under the null hypothesis (g) is given by the number of variables described by the vector autoregression (n) minus the number of cointegrating relations under the null hypothesis (h). In each case the alternative is that there are $h+1$ cointegrating relations.

Source: Michael Osterwald-Lenum (1992), A Note with Quantiles of the Asymptotic Distribution of the Maximum Likelihood Cointegration Rank Test Statistics, Oxford Bulletin of Economics and Statistics 54: p. 462.

Appendix B: Summary Tables

Table B1: *Percentage of profitable and unprofitable trades with closing trigger:* $z_t = 0$

BPM_Ubi

Strategy	*Opening trigger*	*Total number of trades*	*% of Profitable trades*	*% of Unprofitable trades*
1 year, 6 months	2 SDs	1	100%	0%
	3 SDs	0	0%	0%
2 years, 6 months	2 SDs	2	50%	50%
	3 SDs	2	50%	50%

BPM_UniCredit

Strategy	*Opening trigger*	*Total number of trades*	*% of Profitable trades*	*% of Unprofitable trades*
1 year, 6 months	2 SDs	Not cointegrated	-	-
	3 SDs	Not cointegrated	-	-
2 years, 6 months	2 SDs	1	100%	0%
	3 SDs	1	100%	0%

BPM_Intesa

Strategy	*Opening trigger*	*Total number of trades*	*% of Profitable trades*	*% of Unprofitable trades*
1 year, 6 months	2 SDs	2	50%	50%
	3 SDs	2	50%	50%
2 years, 6 months	2 SDs	2	50%	50%
	3 SDs	2	50%	50%

Intesa_Ubi

Strategy	*Opening trigger*	*Total number of trades*	*% of Profitable trades*	*% of Unprofitable trades*
1 year, 6 months	2 SDs	Not cointegrated	-	-
	3 SDs	Not cointegrated	-	-
2 years, 6 months	2 SDs	4	100%	0%
	3 SDs	0	0%	0%

Intesa_UniCredit

Strategy	*Opening trigger*	*Total number of trades*	*% of Profitable trades*	*% of Unprofitable trades*
1 year, 6 months	2 SDs	4	75%	25%
	3 SDs	4	100%	0%
2 years, 6 months	2 SDs	3	100%	0%
	3 SDs	2	100%	0%

Ubi_BPM

Strategy	***Opening trigger***	***Total number of trades***	***% of Profitable trades***	***% of Unprofitable trades***
1 year, 6 months	2 SDs	Not cointegrated	-	-
	3 SDs	Not cointegrated	-	-
2 years, 6 months	2 SDs	1	0%	100%
	3 SDs	1	0%	100%

Ubi_Intesa

Strategy	***Opening trigger***	***Total number of trades***	***% of Profitable trades***	***% of Unprofitable trades***
1 year, 6 months	2 SDs	1	100%	0%
	3 SDs	0	0%	0%
2 years, 6 months	2 SDs	2	100%	0%
	3 SDs	1	100%	0%

Ubi_UniCredit

Strategy	***Opening trigger***	***Total number of trades***	***% of Profitable trades***	***% of Unprofitable trades***
1 year, 6 months	2 SDs	2	100%	0%
	3 SDs	2	100%	0%
2 years, 6 months	2 SDs	4	100%	0%
	3 SDs	1	100%	0%

UniCredit_BPM

Strategy	***Opening trigger***	***Total number of trades***	***% of Profitable trades***	***% of Unprofitable trades***
1 year, 6 months	2 SDs	Not cointegrated	-	-
	3 SDs	Not cointegrated	-	-
2 years, 6 months	2 SDs	2	50%	50%
	3 SDs	1	100%	0%

Intesa_BPM

Strategy	***Opening trigger***	***Total number of trades***	***% of Profitable trades***	***% of Unprofitable trades***
1 year, 6 months	2 SDs	3	66.67%	33.33%
	3 SDs	2	50%	50%
2 years, 6 months	2 SDs	1	0%	100%
	3 SDs	1	0%	100%

UniCredit_Intesa

Strategy	***Opening trigger***	***Total number of trades***	***% of Profitable trades***	***% of Unprofitable trades***
1 year, 6 months	2 SDs	4	75%	25%
	3 SDs	4	100%	0%
2 years, 6 months	2 SDs	4	100%	0%
	3 SDs	2	100%	0%

UniCredit_Ubi

Strategy	***Opening trigger***	***Total number of trades***	***% of Profitable trades***	***% of Unprofitable trades***
1 year, 6 months	2 SDs	1	0%	100%
	3 SDs	1	0%	100%
2 years, 6 months	2 SDs	3	100%	0%
	3 SDs	1	100%	0%

In this work for each pair under analysis both cointegration relationship with either one bank or the other as dependent variable is conducted. For each pair reported in the table, the first bank represents the time series that is used as dependent variable during the Engle-Granger test for cointegration, while the second bank represents the independent variable

Source: *Thomson Reuters Eikon Database.*

Table B2: *Percentage of profitable and unprofitable trades with closing trigger:* $z_t > -q$ *and* $z_t < q$*, with* $q = 2,3$

BPM_Ubi

Strategy	***Opening trigger***	***Total number of trades***	***% of Profitable trades***	***% of Unprofitable trades***
1 year, 6 months	2 SDs	1	100%	0%
	3 SDs	0	0%	0%
2 years, 6 months	2 SDs	5	80%	20%
	3 SDs	9	88.89%	11.11%

BPM_UniCredit

Strategy	***Opening trigger***	***Total number of trades***	***% of Profitable trades***	***% of Unprofitable trades***
1 year, 6 months	2 SDs	Not cointegrated	-	-
	3 SDs	Not cointegrated	-	-
2 years, 6 months	2 SDs	4	100%	0%
	3 SDs	2	100%	0%

BPM_Intesa

Strategy	***Opening trigger***	***Total number of trades***	***% of Profitable trades***	***% of Unprofitable trades***
1 year, 6 months	2 SDs	3	66.67%	33.33%
	3 SDs	3	66.67%	33.33%
2 years, 6 months	2 SDs	5	80%	20%
	3 SDs	5	80%	20%

Intesa_Ubi

Strategy	***Opening trigger***	***Total number of trades***	***% of Profitable trades***	***% of Unprofitable trades***
1 year, 6 months	2 SDs	Not cointegrated	-	-
	3 SDs	Not cointegrated	-	-
2 years, 6 months	2 SDs	7	100%	0%
	3 SDs	0	0%	0%

Intesa_UniCredit

Strategy	***Opening trigger***	***Total number of trades***	***% of Profitable trades***	***% of Unprofitable trades***
1 year, 6 months	2 SDs	14	92.86%	7.14%
	3 SDs	17	100%	0%
2 years, 6 months	2 SDs	8	100%	0%
	3 SDs	2	100%	0%

Ubi_BPM

Strategy	***Opening trigger***	***Total number of trades***	***% of Profitable trades***	***% of Unprofitable trades***
1 year, 6 months	2 SDs	Not cointegrated	-	-
	3 SDs	Not cointegrated	-	-
2 years, 6 months	2 SDs	4	75%	25%
	3 SDs	4	75%	25%

Ubi_Intesa

Strategy	***Opening trigger***	***Total number of trades***	***% of Profitable trades***	***% of Unprofitable trades***
1 year, 6 months	2 SDs	5	100%	0%
	3 SDs	0	0%	0%
2 years, 6 months	2 SDs	12	100%	0%
	3 SDs	2	100%	0%

Ubi_UniCredit

Strategy	***Opening trigger***	***Total number of trades***	***% of Profitable trades***	***% of Unprofitable trades***
1 year, 6 months	2 SDs	8	100%	0%
	3 SDs	2	100%	0%
2 years, 6 months	2 SDs	12	100%	0%
	3 SDs	3	100%	0%

UniCredit_BPM

Strategy	***Opening trigger***	***Total number of trades***	***% of Profitable trades***	***% of Unprofitable trades***
1 year, 6 months	2 SDs	Not cointegrated	-	-
	3 SDs	Not cointegrated	-	-
2 years, 6 months	2 SDs	6	83.33%	16.67%
	3 SDs	2	100%	0%

Intesa_BPM

Strategy	***Opening trigger***	***Total number of trades***	***% of Profitable trades***	***% of Unprofitable trades***
1 year, 6 months	2 SDs	4	75%	25%
	3 SDs	2	50%	50%
2 years, 6 months	2 SDs	2	50%	50%
	3 SDs	5	80%	20%

UniCredit_Intesa

Strategy	***Opening trigger***	***Total number of trades***	***% of Profitable trades***	***% of Unprofitable trades***
1 year, 6 months	2 SDs	25	96%	4%
	3 SDs	15	100%	0%
2 years, 6 months	2 SDs	17	100%	0%
	3 SDs	3	100%	0%

UniCredit_Ubi

Strategy	***Opening trigger***	***Total number of trades***	***% of Profitable trades***	***% of Unprofitable trades***
1 year, 6 months	2 SDs	2	50%	50%
	3 SDs	3	66.67%	33.33%
2 years, 6 months	2 SDs	10	100%	0%
	3 SDs	3	100%	0%

In this work for each pair under analysis both cointegration relationship with either one bank or the other as dependent variable is conducted. For each pair reported in the table, the first bank represents the time series that is used as dependent variable during the Engle-Granger test for cointegration, while the second bank represents the independent variable

Source**:** *Thomson Reuters Eikon Database.*

Table B3: *Percentage of profitable and unprofitable trades for the parametrization 2-year formation period and 2-year trading period with closing trigger:* $z_t = 0$

BPM_Ubi

Strategy	***Opening trigger***	***Total number of trades***	***% of Profitable trades***	***% of Unprofitable trades***
2 years, 2 years	2 SDs	1	100%	0%
	3 SDs	1	100%	0%

BPM_UniCredit

Strategy	***Opening trigger***	***Total number of trades***	***% of Profitable trades***	***% of Unprofitable trades***
2 years, 2 years	2 SDs	Not cointegrated	-	-
	3 SDs	Not cointegrated	-	-

BPM_Intesa

Strategy	***Opening trigger***	***Total number of trades***	***% of Profitable trades***	***% of Unprofitable trades***
2 years, 2 years	2 SDs	1	100%	0%
	3 SDs	1	100%	0%

Intesa_Ubi

Strategy	***Opening trigger***	***Total number of trades***	***% of Profitable trades***	***% of Unprofitable trades***
2 years, 2 years	2 SDs	2	100%	0%
	3 SDs	1	100%	0%

Intesa_UniCredit

Strategy	***Opening trigger***	***Total number of trades***	***% of Profitable trades***	***% of Unprofitable trades***
1 year, 6 mont2 years, 2 years hs	2 SDs	1	100%	0%
	3 SDs	1	100%	0%

Ubi_BPM

Strategy	***Opening trigger***	***Total number of trades***	***% of Profitable trades***	***% of Unprofitable trades***
2 years, 2 years	2 SDs	1	100%	0%
	3 SDs	1	100%	0%

Ubi_Intesa

Strategy	***Opening trigger***	***Total number of trades***	***% of Profitable trades***	***% of Unprofitable trades***
2 years, 2 years	2 SDs	Not cointegrated	-	-
	3 SDs	Not cointegrated	-	-

Ubi_UniCredit

Strategy	***Opening trigger***	***Total number of trades***	***% of Profitable trades***	***% of Unprofitable trades***
2 years, 2 years	2 SDs	2	100%	0%
	3 SDs	1	100%	0%

UniCredit_BPM

Strategy	***Opening trigger***	***Total number of trades***	***% of Profitable trades***	***% of Unprofitable trades***
2 years, 2 years	2 SDs	1	100%	0%
	3 SDs	1	100%	0%

Intesa_BPM

Strategy	***Opening trigger***	***Total number of trades***	***% of Profitable trades***	***% of Unprofitable trades***
2 years, 2 years	2 SDs	1	100%	0%
	3 SDs	1	100%	0%

UniCredit_Intesa

Strategy	***Opening trigger***	***Total number of trades***	***% of Profitable trades***	***% of Unprofitable trades***
2 years, 2 years	2 SDs	1	100%	0%
	3 SDs	1	100%	0%

UniCredit_Ubi

Strategy	***Opening trigger***	***Total number of trades***	***% of Profitable trades***	***% of Unprofitable trades***
2 years, 2 years	2 SDs	1	100%	0%
	3 SDs	2	100%	0%

In this work for each pair under analysis both cointegration relationship with either one bank or the other as dependent variable is conducted. For each pair reported in the table, the first bank represents the time series that is used as dependent variable during the Engle-Granger test for cointegration, while the second bank represents the independent variable

Source: *Thomson Reuters Eikon Database.*

Table B4: *Percentage of profitable and unprofitable trades for the parametrization 2-year formation period and 2-year trading period with closing trigger:* $z_t > -q$ *and* $z_t < q$, *with* $q = 2{,}3$

BPM_Ubi

Strategy	***Opening trigger***	***Total number of trades***	***% of Profitable trades***	***% of Unprofitable trades***
2 years, 2 years	2 SDs	4	100%	0%
	3 SDs	10	100%	0%

BPM_UniCredit

Strategy	***Opening trigger***	***Total number of trades***	***% of Profitable trades***	***% of Unprofitable trades***
2 years, 2 years	2 SDs	Not cointegrated	-	-
	3 SDs	Not cointegrated	-	-

BPM_Intesa

Strategy	***Opening trigger***	***Total number of trades***	***% of Profitable trades***	***% of Unprofitable trades***
2 years, 2 years	2 SDs	3	100%	0%
	3 SDs	10	100%	0%

Intesa_Ubi

Strategy	***Opening trigger***	***Total number of trades***	***% of Profitable trades***	***% of Unprofitable trades***
2 years, 2 years	2 SDs	11	100%	0%
	3 SDs	2	100%	0%

Intesa_UniCredit

Strategy	***Opening trigger***	***Total number of trades***	***% of Profitable trades***	***% of Unprofitable trades***
1 year, 6 mont2 years, 2 years hs	2 SDs	7	100%	0%
	3 SDs	5	100%	0%

Ubi_BPM

Strategy	***Opening trigger***	***Total number of trades***	***% of Profitable trades***	***% of Unprofitable trades***
2 years, 2 years	2 SDs	7	100%	0%
	3 SDs	10	100%	0%

Ubi_Intesa

Strategy	***Opening trigger***	***Total number of trades***	***% of Profitable trades***	***% of Unprofitable trades***
2 years, 2 years	2 SDs	Not cointegrated	-	-
	3 SDs	Not cointegrated	-	-

Ubi_UniCredit

Strategy	*Opening trigger*	*Total number of trades*	*% of Profitable trades*	*% of Unprofitable trades*
2 years, 2 years	2 SDs	14	100%	0%
	3 SDs	1	100%	0%

UniCredit_BPM

Strategy	*Opening trigger*	*Total number of trades*	*% of Profitable trades*	*% of Unprofitable trades*
2 years, 2 years	2 SDs	7	100%	0%
	3 SDs	19	100%	0%

Intesa_BPM

Strategy	*Opening trigger*	*Total number of trades*	*% of Profitable trades*	*% of Unprofitable trades*
2 years, 2 years	2 SDs	4	100%	0%
	3 SDs	9	100%	0%

UniCredit_Intesa

Strategy	*Opening trigger*	*Total number of trades*	*% of Profitable trades*	*% of Unprofitable trades*
2 years, 2 years	2 SDs	8	100%	0%
	3 SDs	7	100%	0%

UniCredit_Ubi

Strategy	*Opening trigger*	*Total number of trades*	*% of Profitable trades*	*% of Unprofitable trades*
2 years, 2 years	2 SDs	18	100%	0%
	3 SDs	2	100%	0%

In this work for each pair under analysis both cointegration relationship with either one bank or the other as dependent variable is conducted. For each pair reported in the table, the first bank represents the time series that is used as dependent variable during the Engle-Granger test for cointegration, while the second bank represents the independent variable

Source: *Thomson Reuters Eikon Database.*

Appendix C: Matlab Code

```
%We present the Matlab code used in this dissertation. Here, we present the
%code only for the pair BPM_Intesa. The same code has been used for all the
%other pairs examined.

clear;
clc;
data=readtable('BPM_Intesa_5years.xlsx');
dates=data(1:end,1);
dates=dates{:,1};
assetNames=data.Properties.VariableNames(2:end);
assetPrice=data(:,assetNames).Variables;
BPM=assetPrice(:,1);
Intesa=assetPrice(:,2);

%Plot of the time series under analysis over the 5-year trading horizon

figure
plot(dates,assetPrice(1:end,1:2),'LineWidth',2)
xlabel('Years')
ylabel('Price')
names=assetNames(1:end)
legend(names,'Location','SW')
title('{Price series}')
axis tight

%Preliminary Analysis: Testing for the Order of (Co-)Integration

%%Unit Root test: Augmented Dickey-Fuller Test
%BPM
[h1,pVal1,stat1,cValue1,reg1]=adftest(BPM,'model','ARD','lags',[0:10]);
[h1D,pValD1,statD1,cValueD1]=adftest(diff(BPM),'model','ARD','lags',[0:10])
;

%Intesa
[h12,pVal2,stat2,cValue2]=adftest(Intesa,'model','ARD','lags',[0:10]);
[h1D2,pVal1D2,statD2,cValueD2]=adftest(diff(Intesa),'model','ARD','lags',[0
:10]);

%%Stationarity test: KPSS test
%BPM
[h_kpss1,pValue_kpss1,statkpss1,cValuekpss1]=
kpsstest(BPM,'trend',false,'lags',[0:10]);
[h_kpss1D,pValue_kpss1D,statkpssD1,cValuekpssD1]=
kpsstest(diff(BPM),'trend',false,'lags',[0:10]);

%Intesa
[h_kpss2,pValue_kpss2,statkpss2,cValuekpss2]=
kpsstest(Intesa,'trend',false,'lags',[0:10]);
[h_kpss2D,pValue_kpss2D,statkpssD2,cValuekpssD2]=
kpsstest(diff(Intesa),'trend',false,'lags',[0:10]);


save BPM_Intesa_5years.mat



%BPM_Intesa_Pairs Trading Strategy

clear
```

```
clc
load BPM_Intesa_5years.mat
Y_d=BPM_Intesa(:,1:2);



%BPM_Intesa_function_5years: this is the function used to produce a trading
signal for our pairs trading strategy

%M represents the length of the formation period
%N represents the length of the trading period
%delta represents the opening trigger
%scaling represents a scaling factor
%cost represents the amount of transaction cost

function [s, reg1, r, res, trn, indicate] = pairs(dates,series, M, N, delta,
scaling, cost)

%% Sweep across the entire time series
%We use the information of the preceding M days to estimate the cointegration
%relationship (if it exists) for the following N days. A new formation period
%is initiated after N days and the procedure is repeated until the end of
the %sample considered.
% We then use this estimated relationship to identify trading opportunities
% until the end of the trading period.

s = zeros(size(series));
indicate = zeros(length(series),1);

for i = M : N : length(s)-N
    % Calibrate cointegration model using the Engle-Granger methodology.
    [h,~,~,~,reg1] = egcitest(series(i-M+1:i, :));
    if h ~= 0

      % Only engage in trading if we reject the null hypothesis that no
      % cointegrating relationship exists, i.e. only if h==0. Conversely, if
      %the time series are found to be not cointegrated, no pairs trading
      %strategy will be implemented in the subsequent trading period.

      % The pairs trading strategy:

      % 1. Compute residuals over next N days
            res     =     series(i:i+N-1,     1)     -     (reg1.coeff(1)     +
            reg1.coeff(2).*series(i:i+N-1, 2));

      % 2. If the residuals are large and positive, then the first series
      %  is likely to decline vs. the second series.   Short the first
      % series by 1 share and long the second series by a scaled number equal
      %to the estimated cointegration coefficient) of shares.   If the
      %residuals are large and negative, do the opposite.

        indicate(i:i+N-1) = res/reg1.RMSE;
        if reg1.RMSE*delta>cost
            j=1;
            s(i,1)=(res(j)/reg1.RMSE < delta)-(res(j)/reg1.RMSE > -delta);
            i=i+1;
            for j=2:N
                if(s(i-1,1)~=0)*(res(j)*res(j-1)>=0)
                    s(i,1)=s(i-1,1);
                else
```

```
                    s(i,1)=(res(j)/reg1.RMSE<delta)-(res(j)/reg1.RMSE>-
delta);
                end
                i=i+1;
            end
            i=i-N;
        end
        s(i:i+N-1, 2) = -reg1.coeff(2) .* s(i:i+N-1, 1);
    end
end

%% Calculate performance statistics: r represents the cumulative profits of
%% our pairs trading strategy

r  = sum([0 0; s(1:end-1, :) .* diff(series) - abs(diff(s))*cost/2] ,2);


% Count the numbers of trades
 trn=0;
 for i=M:length(series)
     if (s(i,1)~=0)*(s(i-1,1)~=s(i,1))
         trn=trn+1;
     end
 end
   %% Plot results

    %% Plot of the time series under analysis

    ax(1) = subplot(3,1,1);
    plot(dates,series,'LineWidth',2),
    grid on
    legend('BPM','Intesa')
    title(['Time Series’)])
    ylabel('Price (€)')

   %% Plot of the normalized spread with the relative opening trigger

    ax(2) = subplot(2,1,1);
    plot(dates,[indicate,delta*ones(size(indicate)),-
delta*ones(size(indicate))])
    grid on
    legend(['Normalized           Spread'],'BPM:           Overvalued','BPM:
Undervalued','Location','NorthWest')
    title(['BPM_Intesa:  Formation  period  =  '  num2str(M)  '  days,  Trading
period = ' num2str(N) ' days, Target Deviation = ', num2str(delta), ' Standard
Deviations'])
    ylabel('Indicator')

%% Plot of the cumulative profits

    ax(3) = subplot(2,1,2);
    plot(dates)
    grid on
    legend('Annualized Returns', 'NorthWest')
    title(['Annualized Returns'])
    ylabel('Returns (%)')
    xlabel('Date')
    linkaxes(ax,'x')
end
```

```
%% Exploiting the function BPM_Intesa_function_5years, we compute the results
%%for different combination of formation period, trading period and target
%%deviations

figure(1)
[s,reg1,r,res,trn] = BPM_Intesa_function_5years(dates,Y_d,252,126,2,252,0);
figure(2)
[s,reg1,r,res,trn] = BPM_Intesa_function_5years (dates,Y_d,252,126,3,252,0);
figure(3)
[s,reg1,r,res,trn] = BPM_Intesa_function_5years (dates,Y_d,504,126,2,252,0);
figure(4)
[s,reg1,r,res,trn] = BPM_Intesa_function_5years (dates,Y_d,504,126,3,252,0);
```