Department of Economics and Finance
Equity markets and alternative investments

Performance Analysis of a Cointegration-Based Statistical Arbitrage Portfolio

Advisor
prof. Paolo Vitale

Co-advisor
prof. Marco Morelli

721221
Federico Cappuccio

Academic Year: 2020/2021

Contents

Chapter 1

Introduction

Algorithmic trading is a rapidly expanding reality. Although it was mainly practiced by banks in the past, more and more retail investors became attracted to it during the years. This phenomenon is, for sure, to be addressed primarily to innovations in the tech industry, along with progress in the A.I. (i.e., artificial intelligence) field, and to the increasing ease of information retrieval for this kind of matter.

In this thesis, we will address the performance analysis of a statistical arbitrage strategy. On top of being particularly profitable, Statistical arbitrage portfolios are also likely to be market neutral, benefiting from both upturns and downturns of the markets. Therefore in periods where markets are extremely volatile (for instance, during Spring 2019, with the Covid-19 pandemic), such strategies should be particularly appealing due to their features. Moreover, because of their intrinsic characteristics, these strategies are ideal to be implemented with an algorithmic tool. Indeed they need very precise price levels to open and close positions along with various statistical tests.

The main objectives of this study are to define the characteristics of a statistical arbitrage strategy and analyze the performance of the associated portfolio through the use of a self-developed algorithm, written with Python, which, relying on sophisticated statistical tools, permits to analyze and evaluate blocks of market data for a broad set of activities. Furthermore, the analysis will be carried on in two different markets in order to highlight the change in the performance according to the different structure and characteristics of the reference market.

After introducing the topic and presenting the main features of the strategy, we will then try to investigate the factors that affect the performance.

In particular, in the second chapter, after digging a brief digression about long-short equity strategies and market neutrality, we will introduce the concept of pair trading before defining statistical arbitrage strategies.

In the third chapter, we will outline all the choices made in structuring the strategy and the algorithm associated with it. After explaining how and which data are selected, we will go through all the statistical tests and calculations needed to generate the trading signals. Once there, we will then thoroughly analyze how returns are computed, also lingering on the needed assumptions.

In the fourth chapter, firstly, we will outline how pairs are traded in each period, and later, after dwelling on the definition of some useful statistical concepts, we will go through the performance analysis in the two different markets. Eventually, after evaluating the market-neutrality of our portfolios, we will consider a risk-adjusted measure, the Sharpe ratio, to better understand the quality of the performance of said portfolios.

Lastly, the fifth chapter will outline the most significant risks that could affect our strategy and then, using the liquidity measure provided by Pastor and Stambaugh, a focus will be made for assessing the exposure to one of the major risks for a statistical arbitrage portfolio: the market liquidity risk.

Chapter 2

Pair trading and statistical arbitrage

In this chapter, firstly, we will focus on giving a brief introduction to different trading strategies, in particular lingering on long/short strategies and market neutrality. Then we will go through a definition of pair trading and statistical arbitrage, highlighting how these strategies are developed and implemented through different approaches.

The general goal of a trading strategy is to maximize profits while also minimizing the risk deriving from investments in the financial marketplaces. Thousands are the strategies that can be implemented to reach this fundamental goal: some traders focus on technical analysis while others on fundamental analysis. With the growing complexity of financial products and increasing volumes of market transactions, more and more sophisticated trading strategies have been implemented in time, and a group of them, often utilized by hedge funds, is the family of long/short equity strategies.

2.1 Long/short equity strategies

Long/short equity strategies are pretty common among hedge funds. Ideally, this family of strategies (many variations are possible) tries to exploit profit opportunities, both from rallies and falls in stock prices. Basically, they aim to open a long position in undervalued stocks and a short position in overvalued ones. Of course, a trader investing with such a strategy would want to profit from both positions in a perfect scenario. However, this is not always possible due to unpredicted market movements or hedging needs (since often short positions are taken in order to hedge the portfolio from the downside risk), but at the end of the day, what matters whenever positions are closed out is that the net profit is positive.

To better understand this kind of strategy, let us consider the following example:

We have two stocks, A and B. Company A is a promising company with a stable financial situation, and its price per share is $p_a = \$40$. B, on the contrary, is a declining company that is being wiped out by competition; its price per share is $p_b = \$30$. A long/short equity trader would want to open a long position in stock A and a short position in stock B. The investor buys 1000 shares of company A at $40 each and 2000 shares of company B at $20 each[1]. If the price of stock A rises to $43 per share and the price of stock B falls to $19 per share, the investor would take advantage of both positions, making $5000 of profit. However, if the company B releases a new patent and its price per share rises to $21 instead of falling, the profit will shrink to $1000: the long position will close with a positive profit of $3000, but the short one will cause a loss of $2000.

2.1.1 Market neutrality

Market neutrality, just like dollar neutrality, is a characteristic that long/short equity strategies can have, and in general, aim to have. In accordance with the CAPM, the β is a measure of risk that tells us the volatility of a stock compared to the volatility of the market. The market has a β equal to one. If a given stock has a beta greater than one, it means that it is more volatile than the market, the contrary otherwise.

A strategy is said to be market-neutral whenever the total exposure to the market expressed in betas over the two positions (long and short) elides each other, i.e., the weighted sum of the betas is zero, where the dimensions of each position give the weights. Suppose an investor is interested in two stocks C and D with the same price, where $\beta_c = 0.5$ and $\beta_d = 2$. If he wants

[1]note that in a situation like this, our investor not only is investing following a long/short strategy, but he is also dollar neutral since the net initial cash flows of the position sum up to zero

to buy 20 shares of stock C, then he will have to short 80 shares of stock D in order to be market neutral, since: $\beta_p = 2 \cdot 20 - 0.5 \cdot 80 = 0$.

Generally, a portfolio is considered to be market-neutral even if its resultant beta is not exactly equal to zero, but it is still a negligible value. If such a condition is verified, in practice, this means that portfolio returns are uncorrelated with market returns. Therefore a market-neutral strategy can generate profits even in a declining or turbulent market.

2.2 Pair trading

A pretty popular variation of the traditional long/short strategy is "pair trading". This term indicates all those strategies that aim to trade "connected" stocks due to similar fundamentals and usually appertaining to the same sector.

In short, a pair trading investor should:

- Identify pairs of stocks that move together, or at least that move in a very similar and "connected" way.
- Open a long position in the undervalued/underpriced stock and a short position in the overvalued/overpriced one whenever the normalized price spread of said stocks diverges enough from the usual pattern. Sooner or later, their prices should return to their equilibrium level.
- Close positions whenever the spread returns to zero, i.e., whenever price charts cross each other.

Even though an investor following these exact same steps, should be in profit by construction with this strategy, these kinds of opportunities are temporary and often automatically and rapidly corrected by the market itself, which pulls prices back to their equilibrium levels. Moreover, the assumptions and beliefs of a given investor may be wrong or not taking into account the big picture in its entirety. For these reasons, any investor trading in accordance with such strategies is not automatically supposed to make profits but, in fact, can still occur into losses from one or both positions, just like explained in the example of paragraph 2.1.

Although the concepts behind the strategy are pretty straightforward and simple to understand, actually finding pairs that move together is not an easy task. As we already said, in the past naive-pair traders would have used fundamental valuation, and particularly financial ratios, in order to find tradable pairs. Nowadays, this method has been set aside in favor of methodologies involving the use of more resilient and reliable statistical relationships.

Whenever pair trading is performed with the help of statistical tools and quantitative analysis, it goes by the name of "statistical arbitrage".

2.3 Statistical arbitrage

In developing pair trading's statistical arbitrage, the first was a Wall Street quant in the mid-1980s: Nunzio Tartaglia. Tartaglia, at that time, gathered a team of academics (mainly composed of computer scientists, physics, and mathematicians) in order to develop an automated trading strategy capable of exploiting arbitrage opportunities and market inefficiencies through the systematic use of quantitative and statistical instruments. Tartaglia and his team performed statistical arbitrage for Morgan Stanley in 1987, generating about 50 million dollars of profits for the firm. As time passed, Tartaglia's strategy gathered increasing consensus, becoming today one of the most popular market-neutral strategies among long/short equity funds. [2]

There are several ways to identify tradable pairs in accordance with statistical arbitrage. One of them, and probably among the most famous and simplest quantitative method, is the one utilizing a measure of distance. Such measure of distance can be provided, for instance, by the tracking variance between normalized prices of the two time series. Supposing to analyze two generic stocks A and B, over a period T that goes from 0 to T, normalized prices are defined as:

$$Q_t^A = \frac{P_t^A}{P_0^A} \tag{2.1}$$

and:

$$Q_t^B = \frac{P_t^B}{P_0^B} \tag{2.2}$$

[2] Gatev et al., 2006

thus, the tracking variance (TV) for the two stocks will be:

$$TV = \frac{1}{T}\sum_{t=0}^{T}(Q_t^A - Q_t^B)^2 \tag{2.3}$$

Since the tracking variance, thus computed, is configured as the estimated average price distance, only pairs with the smallest tracking variance will be selected. Once pairs are selected, we need to define the rules for opening and closing positions. For example, Gatev et al. (2006) utilize as a threshold to trigger position opening, a value equal to twice the standard deviation of the normalized spread of the two stocks (σ):

$$\sigma = (\frac{1}{T-1}\sum_{t=0}^{T}[\Delta_t - \bar{\Delta}]^2)^{\frac{1}{2}} \tag{2.4}$$

Whenever the absolute value of the spread $\Delta_t = Q_t^A - Q_t^B$ is greater than $2 \cdot \sigma$ a position is opened and in particular:

- A long position in stock A and a short position in stock B are opened when $\Delta_t > 2 \cdot \sigma$
- A short position in stock A and a long position in stock B are opened when $\Delta_t < -2 \cdot \sigma$

Then positions are closed out whenever the spread Δ returns to zero or it becomes big enough to hit a possible set stop-loss.

Despite the fact that this strategy may be profitable, the tracking variance may not be a reliable measure to identify pairs. In order to overcome this issue over the years, many different models have been developed. In this work, we will focus on the approach that makes use of the cointegration, an approach that can grant more thorough and reliable results, as we will see later in chapter 3.5.2.

As we said, there are many ways to identify pairs for such strategies. For instance, we could address the Johansen cointegration approach, which will lead to trade sets (i.e., triplets, quadruplets, and so on) of stocks instead of pairs, or again, machine learning techniques used to rank different pairs. However, these will not be treated in this dissertation.

Chapter 3

Trading strategy

In this chapter, we are going to discuss and explain the choices made in structuring the strategy that led us to the construction of our statistical arbitrage-based portfolio. Starting from how a strategy is tested and validated and how data have been chosen, we will gradually pass through all the steps that will lead us to calculate returns and, consequently, measure the portfolio's performance. These steps can be summarized as follows:

1. Data picking and sampling
2. Data polishing
3. Pair selection
4. Trading signal generation
5. Portfolio returns

3.1 Data picking and sampling

3.1.1 Strategy validation

Before talking about the actual steps involved in the strategy, it may be helpful to make a brief digression to clarify the different analyses needed to test the validity of a given trading strategy. In order to validate a strategy, it is necessary to test its performance over two different kinds of analysis: in-sample and out-of-sample analyses. The former is carried out to test the validity of a strategy and, in particular, to see if the strategy is profitable when certain given statistical relationships are met. The latter is done to evaluate the actual performance of the strategy. We need to distinguish between these two analyses since, with the in-sample one, we are testing the strategy over a period in which we already know such statistical relationships do exist; it is like we have information from the future. The performance will clearly be biased, so we can only use it to check if the strategy is valid or not: a robust and reliable strategy will have abnormal returns [1]; on the contrary if the strategy lacks robustness, returns will be poor. Coming to the out-of-sample analysis, this is the one that actually measures the strategy's performance by testing it out "live" with no information from the future and assuming that relationships found and existing in the previous periods will persist in the current one. If this assumption is valid, usually, the strategy will be profitable.

3.1.2 Data and sampling

The study involved daily price data from the first quarter of 2011 (2011Q1) to the second quarter of 2021 (2021Q2) for stock that are part of two major indices: SP500 for the United States market, and Euro Stoxx 50 for the European one. All data were selected from the Refinitiv databases. The last ten years of data were chosen and sampled into subsets to provide a thorough and solid analysis. On the one hand, since we want the in-sample statistical relationships to be strong enough to persist in the out-of-sample period (i.e., trading period), they must be tested over a significant period of time. On the other hand, however, there is a trade-off between the robustness of statistical relationships and sensitivity of said relationships to new shocks: the bigger the sample, the lesser and the slower it will react to further shocks. Hence the related in-sample dataset has to be big enough to reflect the first need but still not so extended to cause a slowdown in reception and absorption of new information. On the other hand, the out-of-sample dataset has to be short enough for us to be sufficiently sure that the statistical relationships will persist throughout the period, but not so short to make the analysis not relevant.

For said reasons, the observations have been split into two initial datasets, one for the in-sample checks (like pair selection) and another for assessing the

[1] even in the order of 100%, or more

out-of-sample performance. After various tests, evidence showed that three years of observations would be the optimal period length for the in-sample dataset since it allowed for both robustness and shock absorption. However, the out-of-sample period had to be chosen of a much shorter length: only six months, to maximize the chance of the relationships persisting and since the end of quarter 2 (Q2) and quarter 4 (Q4) are critical deadlines for companies and in the financial calendar in general. Once all of the analyses on the two datasets (which merged together include three years and six months of observations) are done, they roll over a time window of six months so that the in-sample dataset drops its oldest two quarters and gains the next most recent semester (the one of the old out-of-sample dataset), and the out-of-sample dataset become composed by the immediately next, most recent semester. The process is repeated repeatedly until the out-of-sample dataset reaches the second quarter of 2021. This methodology allows us to capitalize on information that companies release every semester, helping us build a much more solid and reliable model.

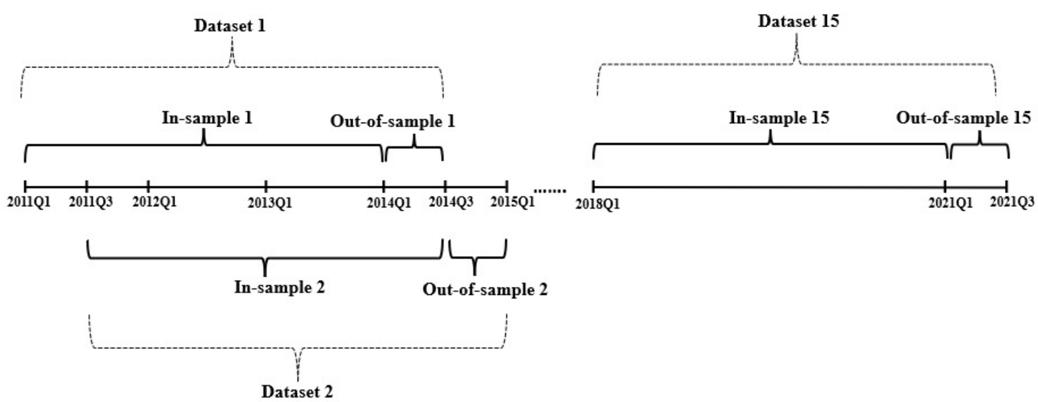


Figure 3.1: Datasets structure

The structure of our datasets is summarized in figure 3.1, and goes as follows: the first in-sample dataset starts on 01/01/2011 and ends on 31/12/2013, while the first out-of-sample dataset starts on 01/01/2014 and ends on 30/06/2014. Going on with the iterating process mentioned above we will finally reach our last dataset, where the in-sample one will start on 01/01/2018 and will end on 31/12/2020, and the out-of-sample one will start on 01/01/2021 and will end on 30/06/2021.

3.2 Data polishing

After being sliced into many in-sample and out-of-sample datasets, according to the methodology mentioned in the previous paragraph, all the time series with any missing values need to be polished. If they have a relevant number of missing values in the considered period, they will be dropped. On the contrary, if missing values are limited to closed market days or a minimal portion of the dataset, they are filled out using the Python command "fillna" and specifying the "ffill" method that will automatically fill any missing value with the one which precedes it.

Once our dataset is polished and we have the same number of observations for every series of the assets, we can proceed to start our analysis.

3.2.1 Stationarity

A stationary time series is one whose properties do not depend on the time at which the series is observed. In the most intuitive sense, a stationary time series is a series where the statistical properties of the process do not change over time. Though this does not mean that the series does not change at all over time, but means that what remains invariant is just the way it changes over time. To better grasp this concept: a linear function could be interpreted as the algebraic equivalent of a stationary time series, not a constant one. The value of a linear function changes as x variates, but the way it changes remains constant; it has a constant slope, there is only one value that captures that rate of change. More formally: a stationary time series is a series whose properties do not depend on the time at which the series is observed. In general, a stationary time series will have no predictable patterns in the long term. Time plots will show the series as roughly horizontal (although some cyclic behaviour is possible), with constant variance[2]. A graphical representation of the process is given in figure 3.2.

In order to better understand a more thorough and formal definition of what a stationary process is, it may be useful to introduce a formal definition of stochastic processes.

[2] Hyndman and Athanasopoulos, 2018, chap. 8.1

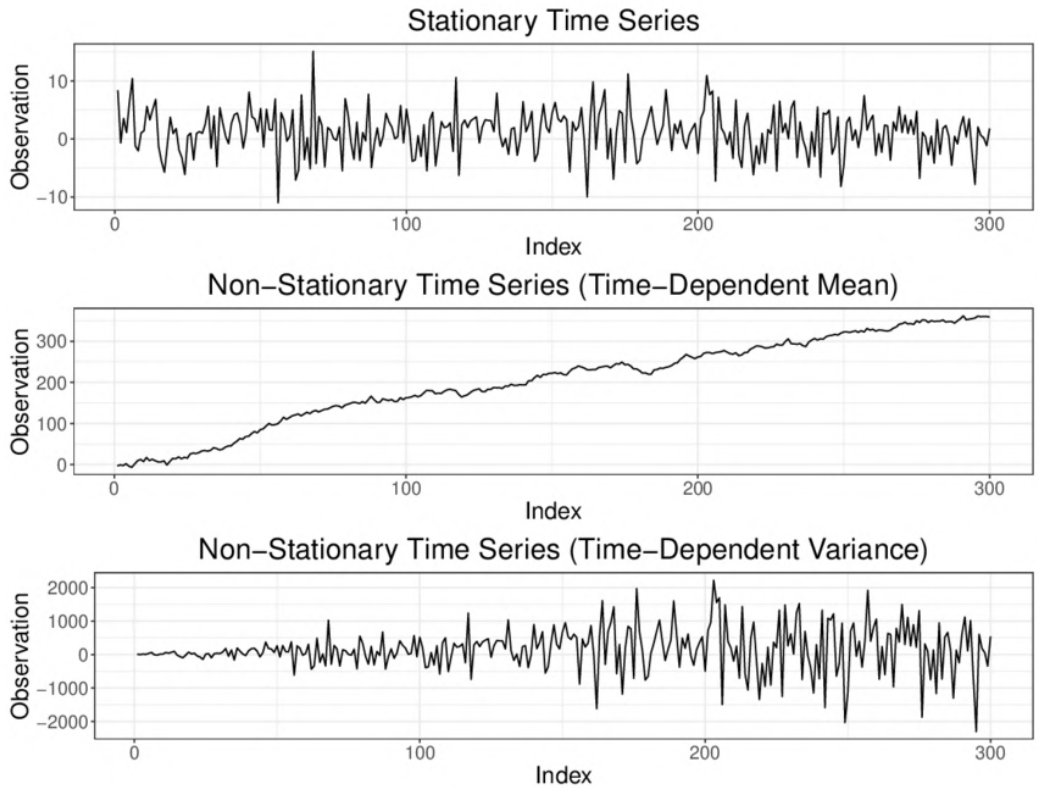


Figure 3.2: Stationary vs. non-stationary time series

Stochastic processes

A stochastic process is a family of real random variables $X = \{x_i(\omega); i \in T\}$ existing and defined all on the same probability space (Ω, F, P), where T is the index set of the process. If $T \subset Z$, then the process will be a discrete stochastic process. If T, instead, is a subset of R, the process is then a continuous stochastic process.

For a finite set of integers $T = \{t_1, \ldots, t_n\}$, so that $T \subset Z$, the cumulative distribution function of $X = \{x_i(\omega); i \in T\}$ is defined as follows:

$$F_{t_1,...,t_n}(x_{t_1}, ..., x_{t_n}) = P(X_{t_1}(\omega)) \leq (X_{t_1}(\omega)), ..., (X_{t_n}(\omega)) \leq (X_{t_n}(\omega)) \quad (3.1)$$

which, for the generic stochastic process X is indicated as:

$$F_X(x_{t_1}, ..., x_{t_n}) \quad (3.2)$$

The finite-dimensional distribution of a stochastic process is then defined as the set of all the joint distribution functions for every of the finite integer sets T of any dimension n. Considering a discrete process it will be the set:

$$\{F_X(x_{t_1}, ..., x_{t_n}); n \in Z^+, T \subset Z\} \tag{3.3}$$

Equation 3.3 substantially represents a projection of the process onto a finite-dimensional vector space (and in particular, since we are dealing with time series,it represents a finite set of points in time).

Now that we have defined a stochastic process, a formal definition of stationarity can be provided. In particular, we will consider what in literature is known as "weak stationarity" or "wide-sense stationarity".

Weak stationarity

As already said, if a series is stationary will mean that the statistical properties of the process do not change over time. Focusing on the weak stationarity, the statistical properties involved are the mean, the variance, and the covariance. So formally, we have that the process $X = \{x_i; i \in Z\}$ is weakly stationary if:

1. $E[x_i] = \mu, \forall t$
2. $E[x_i^2] < \infty, \forall t$ and therefore $E[(x_i - \mu)^2] < \infty, \forall t$
3. $cov(x_u, x_v) = cov(x_{u+a}, x_{v+a}), \forall u, v, a$

The first condition means that the first moment (i.e. the mean) of the process is constant for every t.
the second says that the second moment (i-e the variance) of the process must be finite for every t.
The third one states that the cross moment (i.e., the covariance) depends only on the difference u-v and implies that every lag $\lambda \in N$ has constant covariance, and, therefore, the process will also have constant variance:

$$cov(X_{t1}, X_{t2}) = K_{XX}(t_1, t_2) = K_{XX}(t_2 - t_1, 0) = K_{XX}(\lambda) \tag{3.4}$$

then:

$$Var(X_t) = cov(X_t, X_t) = K_{XX}(t, t) = K_{XX}(0) = d \tag{3.5}$$

So weak stationary processes are characterized by constant mean and variance.

Testing for stationarity: the Augmented Dickey-Fuller test

Since this strategy needs a test of cointegration (Engle-Granger test) to evaluate statistical relationships between stocks, originally stationary series cannot be taken into account, as we will see later in this chapter. Thus, for said reasons, stationary series are detected and then dropped from the in-sample dataset.

In order to detect any stationary series, the augmented Dickey-Fuller test comes in help. This test fundamentally is a statistical significance test that can be used for testing the existence of a unit root in a univariate process in the presence of serial correlation. The null hypothesis of the test asserts that a unit root does exist and, therefore, that the series is not likely to be stationary. The alternative hypothesis is that there is no unit root, and so the series is likely to be stationary[3]. If the p-value is above a critical value, then we cannot reject the null hypothesis, and so there is a unit root. Alternatively, the null hypothesis is rejected in favor of the alternative one. Discussing it formally:

Given an AR(p) model and the characteristic polynomial associated with it:

$$y_t = \theta_1 y_{t-1} + \theta_2 y_{t-2} + ... + \theta_p y_{t-p} + \epsilon_t \tag{3.6}$$

$$\theta(L) = 1 - \theta_1 L - \theta_2 L^2 - ..... - \theta_p L^p \tag{3.7}$$

where L is the lag operator and is defined as $Ly_t = y_{t_{-1}}$.
If a unit root exists, this will mean that the characteristic polynomial evaluated in one, is equal to zero:

$$\theta(1) = 0 \tag{3.8}$$

which is equivalent to say:

$$1 - \theta_1 - \theta_2 - ..... - \theta_p = 0 \tag{3.9}$$

As said before, the Augmented Dickey Fuller aims to test whether a unit root exists or not checking if 3.9 equation has an existing solution under the hypothesis:

[3] note that since this is a statistical significance test, it is not guaranteed that the series will be or not stationary. The accuracy of the result is strictly linked to the confidence level selected for the p-value

1. H_0 : there is a solution to the 3.9 (i.e., there is a unit root)
2. H_1 : there is no solution to the 3.9 (i.e., there is no unit root)

if the null hypothesis is accepted, the series will likely be non-stationary. On the contrary, the series will be stationary if it is rejected in favor of the alternative one.

In this work, all the series in the in-sample datasets are tested with the Augmented Dickey-Fuller test before doing any other analysis, using a for loop. Whenever a series is found to be stationary, it is added to a list. Eventually, the list with all the stationary series is dropped from the in-sample datasets to avoid biases in later tests.

3.3 Pair selection

Once datasets are polished and ready to be used, we want to find a certain number of pairs of stocks, so that each member of a single pair is linked one to the other by a meaningful statistical relationship: the cointegration.

3.3.1 Cointegration

Before cointegration tests were introduced, economists used to rely on linear regressions to find relationships between different time series. In the 1980s, Engle and Granger began to claim that the linear regression approach to the analysis of time series was flawed since it could potentially lead to the so-called spurious correlations[4]. Then in 1987, they coined the term and formalized the concept of cointegration for the first time.

Given two time series y_t and s_t, known to be non-stationary processes, as in figure 3.3, if y_t and s_t are both I(1) and exists a vector β such that:

$$\beta' x_t = \begin{bmatrix} \beta_1 & \beta_2 \end{bmatrix} \begin{bmatrix} y_t \\ s_t \end{bmatrix} = \beta_1 y_t + \beta_2 s_t \sim I(0) \tag{3.10}$$

then y_t and s_y are cointegrated with β as their cointegration vector[5].

[4] spurious correlations occur whenever two or more variables result to be correlated due to coincidence or unknown factors, conducing to misleading results.

[5] $x_t = \begin{bmatrix} y_t \\ s_t \end{bmatrix}$

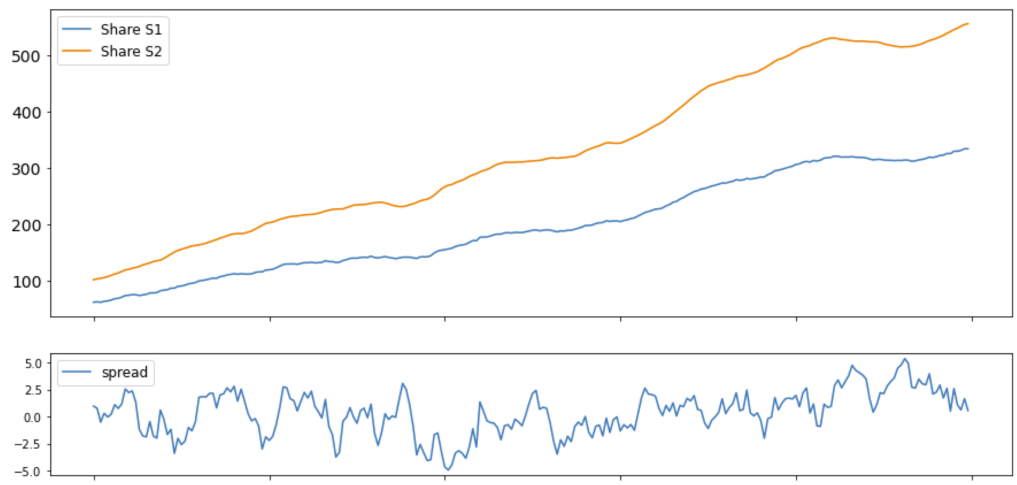


Figure 3.3: Cointegrated series

So, if two time series are integrated of order one and exists a vector beta such that a linear combination of the two time series generates a stationary process, it can be asserted that the two series are cointegrated. In other words, cointegration helps us identify scenarios where two or more non-stationary series are integrated in a way that does not permit them to deviate from the equilibrium in the long run.

Cointegration test

Many tests can be used in order to assess cointegration between two time series. In this work, we utilize the Engle-Granger test, where the null hypothesis is that there is no cointegration. This method is based on a two-step procedure:

1. performs a static regression of the time series, thus generating the residuals

$$y_t = \mu + \beta_2 s_t + u_t \tag{3.11}$$

   where u_t are the residuals of the OLS.

2. tests the stationarity of the residuals checking if a unit root exists by performing an Augmented Dickey-Fuller test.

If there actually is cointegration between the two series, the residuals will be stationary (i.e. no unit root is found in the ADF test). If a unit root is found in this way and the null hypothesis of the ADF test is accepted,and then there is no cointegration. Contrariwise, if there is no unit root and ADF's H_o is rejected, the two series are cointegrated.

3.3.2 Pair ranking

According to a looping methodology similar to the one utilized for testing stationarity of the series in datasets, the Engle-Granger test is performed on any possible combination of series pairwise. This looping test will lead (especially when treating large indices like SP500) to a vast number of cointegrated pairs. Thus the problem is how to rank and pick them in order to maximize the performance, basing on the reliability of the relationship.

Initially, they were ranked based on their p-value. However, this method turned out to be inconsistent and unreliable in assessing the strength of cointegration among pairs.

Eventually, a Half-life test has been added in order to rank pairs accordingly to their mean reversion capability. This method proved to be effective in ranking pairs and actually improved performances. The half-life is defined as the number of periods required for a time series to halve its value, and is calculated as:

$$h = -\frac{log(2)}{log(\rho_1)} \tag{3.12}$$

Where ρ_1 is the autocorrelation of the series at lag one.

3.4 Trading signal generation

In order to define the size of long and short positions within each pair, an OLS regression is run, whose beta will tell us how much the values of a given series of stock change when its cointegrated counterpart changes of one unit. All the betas are then stored, and the preliminary analysis is over.

Moving to the out-of-sample analysis, and in particular moving from the in-sample dataset to the out-of-sample one, for every pair[6], the spread of the

[6]each composed by two generic stock Y and X

two time series is calculated:

$$\Delta(Y_t, X_t) = y_t - x_t \cdot \beta \tag{3.13}$$

and then it is standardized to grant that all data have the same scale. In the end, a "spread matrix" containing the standardized time series of the spreads for every pair is generated.

All these steps will eventually lead to constructing a null matrix (which we will call "signal matrix") with the same dimensions of the spread matrix. This matrix will be used to open and close positions in accordance with the strategy, as the value of the spread exceeds some given thresholds.
The threshold values of the spread that will trigger the opening and closing of a position are, respectively, ± 1 and 0. Moreover, in order to let the algorithm recognize whether it is dealing with a long, a short, or opening a position, a cumulative sum is added to the process so that if the sum turns out to be 0, it will mean that there are no opened positions. Otherwise, there will be a long position (if the sum is equal to +1) or a short position (if the sum is equal to -1). Therefore, we have four cases, two when opening positions and two whenever we are closing them, and the process goes as follows:

1. opening a long position: whenever, at a specific date, the value of a given spread is greater than +1*standard deviation[7], and the associated value of the cumulative sum of previous values in the signal matrix is 0, the value of the signal matrix corresponding to that exact date is changed from 0 to +1.

2. opening a short position: whenever, at a specific date, the value of a given spread is smaller than -1*standard deviation, and the associated value of the cumulative sum of previous values in the signal matrix is 0, the corresponding value of the signal matrix is changed to -1.

3. closing a long position: whenever, at a specific date, the value of a given spread is equal to 0, and the associated value of the cumulative sum of previous values in the signal matrix is +1, the corresponding value of the signal matrix is changed to -1.

[7] Note that since the series is standardized its standard deviation will always be equal to 1

4. closing a short position: whenever, at a specific date, the value of a given spread is equal to 0, and the associated value of the cumulative sum of previous values in the signal matrix is -1, the corresponding value of the signal matrix is changed to +1.

3.4.1 Negative betas

It is worth pointing out that occasionally the OLS regression will output negative betas. In a statistical arbitrage strategy, traded pairs typically have positive betas since they would immediately and intuitively conduct to the construction of a classical long/short strategy. However, in this work, pairs with negative betas are not dropped because doing so would generate a biasing effect due to the fact that pairs are ranked following the criteria mentioned in the paragraph 3.3.2

If a stock Y and a stock X are related by a positive beta (which means that if the price of Y changes by one unit, then the price of X will change of B units in the same direction), we would open a long (or short) position of dimension 1 in stock Y and contextually, a short (or long) position of dimension β in stock X as usual and just like explained in previous sections. Although, when we are dealing with negative betas, it means that stock Y and stock X are negatively correlated, so whenever the price of Stock Y changes by 1 unit, the price of X will change of β units, but this time in the opposite direction. Thus when dealing with negative betas and a long (or short) position is opened in stock Y, the corresponding position in stock X will no more be a short one, but a long (or short) as well.

3.5 Portfolio returns

3.5.1 Assumptions

Before talking about how returns are treated in this work, we need to describe some assumptions that this model makes:

1. Full reinvestment of returns

2. Equally weighted portfolio at the beginning of every period

3. Finite budget

4. Perfect divisibility of asset prices

1. Full reinvestment of returns

Gains made both after each trade and at the end of each period are supposed to be totally reinvested, leading to a continuously compounded computation of returns.

2. Equally weighted portfolio at the beginning of every period

At the beginning of every period (i.e., every semester), our statistical arbitrage portfolio will have equal weightings for every couple considered. As time passes and trades are closed, since capital gains are fully reinvested, every single asset performance will change its own relative weighting, making it heavier (if it performed well) or lighter (if, on the contrary, it performed badly) in regard to the entire portfolio weight in that specific period, in accordance with returns calculated by Gatev as we will see in paragraph 3.5.2. However, no matter how different assets performed in a given semester, the portfolio is systematically rebalanced every six months in order to re-make it equally weighted.

3. Finite budget

In order to bring the analysis closer to reality, we will assume that our budget is not infinite and represented by a generic amount M, which will be allocated among the pairs in equal parts, corresponding to $M\frac{1}{n}$, where n is the number of considered pairs. This amount will then be split among the two assets composing each pair in a way that we will see later on.

4. Perfect divisibility of asset prices

Since assumption three gives us restrictions about our budget, we may not be able to buy the exact quantity that we want of a given stock. Therefore we may be able to buy only a portion of it according to our budget constrain. For this reason, we will assume that any fraction of all treated stocks can be purchased. Even though this assumption may sound irrealistic when utilizing actual stocks, it can be easily implemented when trading with CFDs [8].[9]

3.5.2 Returns computation

Since pairs can open and close positions several times across a single trading period, we will have multiple cash flows for each of them. We will analyze the stand-alone performance of every single asset, and then we will add weights and constraints imposed by our assumptions, to eventually compute portfolio returns.

According to returns calculated by Gatev et al. (2006), returns of the aggregate are calculated as:

$$r_{P,t} = \frac{\sum_{i \in P} w_{i,t} \cdot r_{i,t}}{\sum_{i \in P} w_{i,t}} \tag{3.14}$$

Where $r_{i,t}$ represent the naive return of the stock i at time t, calculated simply by computing its percentage change, and w are the wealth levels, which again:

$$w_{i,t} = w_{i,t-1}(1 + r_{i,t-1}) = (1 + r_{i,1}).....(1 + r_{i,t-1}) \tag{3.15}$$

Equations (3.14) and (3.15) tell us what we mentioned before with the second assumption in paragraph 3.5.1: returns of every asset are affected by the relative weight they gain throughout the trading period due to their performance.

Considering now the allocation of the budget within each pair, and assuming that Y and X are the sieries of the components of a generic pair, the Gatev return (i.e., eq. 3.14) of each asset has to be weighted for the internal allocation ratio (Q) between the two components of each pair. Where for the generic component Y:

$$Q_{Y,t} = \frac{p_{y,t}}{p_{y,t} + |\beta|\, p_{x,t}} \cdot \frac{1}{n} \tag{3.16}$$

[8]Contracts for differences (CFDs) are contracts between an investor and a financial institution, in which the investor opens a position on the future value of an asset. The difference between the open and closing prices of the trade is cash-settled. CFDs are available for a wide range of assets, such as commodities, shares, and many others

[9]cfr. Investopedia, "An Introduction to Contract for Differences (CFDs)"

and for the component X:

$$Q_{X,t} = \frac{|\beta|\, p_{x,t}}{p_{y,t} + |\beta|\, p_{x,t}} \cdot \frac{1}{n} \tag{3.17}$$

Where n is the number of pairs that we are trading.

Note that: $0 < Q < 1$, $Q_Y = 1 - Q_X$ and we are using $|\beta|$ because its sign is only used to determine whether we have to open a long or a short position in stock X. Leaving M undefined and applying 3.16 and 3.17 weightings to the Gatev returns with the same daily marking-to-market procedure, our formula for the return of a single asset with regar to the entire portfolio becomes:

$$r_{p_{i,t}} = \frac{Q_{i,t} \cdot r_{i,t} \cdot w_{i,t}}{\sum_{i \in P} Q_{i,t} \cdot w_{i,t}} \tag{3.18}$$

Where $r_{i,t}$ represent the naive return of the stock i at time t, calculated simply by computing its percentage change And the total return of the portfolio at time t:

$$r_{P,t} = \sum_{i \in P} r_{p_{i,t}} \tag{3.19}$$

Chapter 4

Performance analysis

In this chapter, we will discuss the performance of our cointegration-based statistical arbitrage portfolios in two different stock markets: the U.S. and the European markets. Starting from the raw returns, we will gradually analyze various measures that also take into account the strategy's actual risk to better understand the performance in all its aspects. Furthermore, we will compare the profitability of the strategy in the two markets, which are characterized by profound differences.

4.1 Traded pairs

As explained in paragraph 3.3, in each period, we trade twenty pairs of stocks. Traded pairs will only remain in our portfolio for six months (exactly one period). Then, for optimization reasons, at the end of every period (i.e., every six months), the portfolio is rebalanced, rerunning the cointegration analysis and possibly finding different cointegrated pairs. So, our set of twenty traded pairs may or may not remain the same in each period, depending on the test results. However, it may be interesting to know at least which pairs are traded in the first trading session of the study along with their betas.

For the U.S. market, in the first period we are trading pairs indicated in the table in figure 4.1 below.

Beta	Y	X
1.4213423636792781	BANK OF NEW YORK MELLON	BANK OF AMERICA
4.082638017108977	TYLER TECHNOLOGIES	MARSH & MCLENNAN
1.4336087976682517	GILEAD SCIENCES	PACKAGING CORP.OF AM.
2.4400286725710387	FLEETCOR TECHNOLOGIES	PACKAGING CORP.OF AM.
1.507638379725726	AMETEK	JOHNSON CONTROLS INTL.
2.1546258294844747	WABTEC	MARSH & MCLENNAN
1.7715479748228162	WALGREENS BOOTS ALLIANCE	MOLSON COORS BEVERAGE COMPANY B
3.0996943542724877	CROWN CASTLE INTL.	EBAY
-0.555876780396888	FIDELITY NAT.INFO.SVS.	NEWMONT
1.5753699878046632	T ROWE PRICE GROUP	NEWELL BRANDS (XSC)
1.9626353782140986	KANSAS CITY SOUTHERN	CHUBB
2.42337038268149	DTE ENERGY	CMS ENERGY
1.304220726783303	PRUDENTIAL FINL.	GENERAL DYNAMICS
2.375801688581621	BIO-RAD LABORATORIES 'A'	CBRE GROUP CLASS A
1.0371055660105708	AMERICAN WATER WORKS	ABBOTT LABORATORIES
1.354509043864819	T ROWE PRICE GROUP	RAYMOND JAMES FINL.
2.1363860490700546	T ROWE PRICE GROUP	MASCO
1.017102566068251	PHILIP MORRIS INTL.	FEDERAL REALTY INV.TST.
1.5530755141301624	LIVE NATION ENTM.	BOSTON SCIENTIFIC
-0.6076306800911557	CH ROBINSON WWD.	FEDERAL REALTY INV.TST.

Figure 4.1: Traded pairs in the first period for the U.S. market, along with their betas

Traded pairs and their respective betas for the European market are indicated in the table in figure 4.2.

Beta	Y	X
1.0346991312863776	L AIR LQE.SC.ANYME. POUR L ETUDE ET L EPXTN.	DEUTSCHE POST (XET)
1.2584209540253926	DAIMLER (XET)	DEUTSCHE BOERSE (XET)
1.2115566979361363	DEUTSCHE BOERSE (XET)	VINCI
2.743799380651007	ASML HOLDING	L AIR LQE.SC.ANYME. POUR L ETUDE ET L EPXTN.
1.0232014508476788	L AIR LQE.SC.ANYME. POUR L ETUDE ET L EPXTN.	INDITEX
2.57219133069614	ANHEUSER-BUSCH INBEV	L AIR LQE.SC.ANYME. POUR L ETUDE ET L EPXTN.
1.8684752066503316	LVMH	INDITEX
5.746714473942655	SIEMENS (XET)	ING GROEP
0.9323792629891148	CRH	KONINKLIJKE AHOLD DELHAIZE
1.5371307210807603	SAFRAN	AMADEUS IT GROUP
1.0099157682487487	BNP PARIBAS	SIEMENS (XET)
2.8833293699705504	L AIR LQE.SC.ANYME. POUR L ETUDE ET L EPXTN.	KONINKLIJKE AHOLD DELHAIZE
0.9828454837077097	LVMH	SAP (XET)
2.0982936883568226	SANOFI	KONE 'B'
2.5903731770994183	ADIDAS (XET)	L AIR LQE.SC.ANYME. POUR L ETUDE ET L EPXTN.
2.920436666697552	BAYER (XET)	L AIR LQE.SC.ANYME. POUR L ETUDE ET L EPXTN.
1.472129905586963	LVMH	KONE 'B'
2.6103095065124045	VIVENDI	BANCO SANTANDER
0.8977059310587934	AMADEUS IT GROUP	L AIR LQE.SC.ANYME. POUR L ETUDE ET L EPXTN.
2.405406761771523	SCHNEIDER ELECTRIC	AXA

Figure 4.2: Traded pairs in the first period for the European market, along with their betas

As we can notice from the figures above, the components of our pairs are not necessarily coming from the same sector as happens for the first pair in figure 4.1, where we have Bank of New York Mellon and Bank of America, both obviously belonging to the banking sector. But, on the contrary, they can belong to different industries like in the last pair of the European market where we have Schneider Electric with Axa. Moreover, the same asset may be cointegrated with more than one other asset, making it possible to have the same stock to compare more than one pair as happens for Air Liquide, which appears eight times in the first period.

4.2 Portfolio statistics

Firstly we will analyze the strategy over the entire trading period of eight years in order to assess the overall performance, focusing on returns (also with regard to the overperformance over the respective market index), their distribution, volatility, portfolio beta, and maximum drawdown for the two different markets.

Regarding how returns are distributed and volatility measures, it may be useful to dig a brief digression in order to introduce these measures: kurtosis and skewness for the former aspect, and maximum drawdown for the latter.

Kurtosis

Kurtosis is a statistical measure that defines how fat the tails of a distribution are and how they differ from the tails of a normal distribution. In other words, kurtosis identifies whether or not the tails of a given distribution contain extreme values. A normal distribution has a kurtosis value of 3, so what matters for us is the excess kurtosis of our distribution relatively to the normal one. Thus, as showed in figure 4.3, we have three different cases:

1. Excess kurtosis $= 0$: the distribution follows the normal one in this case, and it is defined as "mesokurtic".

2. Excess kurtosis < 0: the distribution is characterized by plain tails, with a low probability for them to contain extreme values. It is called "platykurtic".

3. Excess kurtosis > 0: the distribution has fat tails, very likely to contain extreme values. In this case, we are speaking about "leptokurtic" distributions.

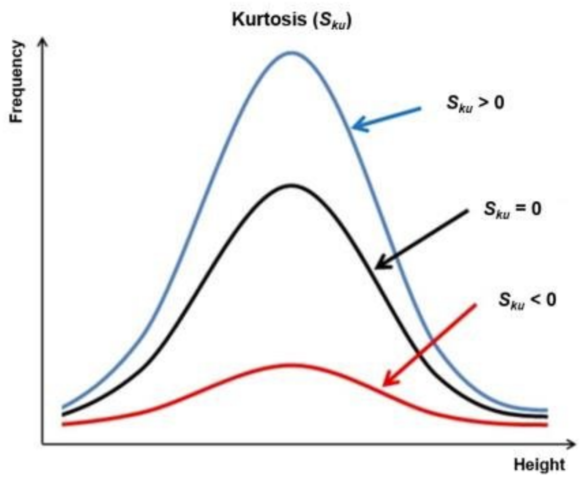


Figure 4.3: Graphical representation of kurtosis.

Skewness

Coming to the skewness, it is a measure of the asymmetry of a distribution. In particular, it measures how much a certain distribution concentrates around its mean, or it disperses either to the right side or to the left one. As for the kurtosis, we will have three cases:

1. Skewness $= 0$: the distribution is symmetrical around its median value, so median, mean, and mode coincide (figure 4.4(b)). This is typical of the normal distribution.

2. Skewness < 0: the distribution is asymmetrical around its median value and presents a concentration of the mass of the distribution on the right side. The mode is greater than the median, and thus the distribution shows a longer left tail (figure 4.4(a)).

3. Skewness > 0: the distribution is asymmetrical around its median value and presents a concentration of the mass of the distribution on the left side. The mode is smaller than the median, and thus the distribution shows a longer right tail (figure 4.4(c)).

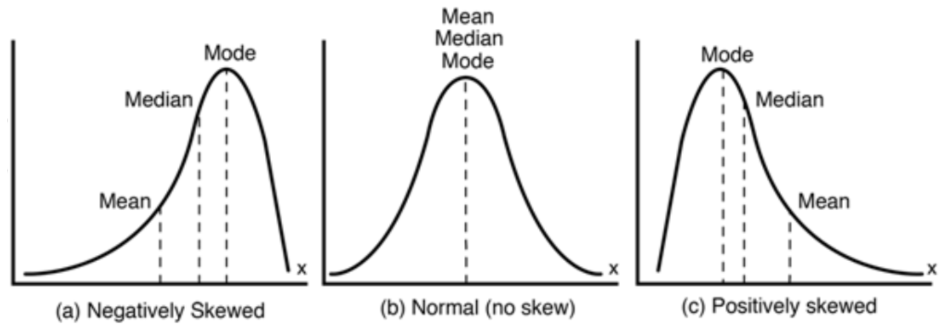


Figure 4.4: Graphical representation of skewness.

Maximum drawdown

Moving on to volatility measures, let us introduce the concept of maximum drawdown. The maximum drawdown (MDD) represents an indicator of downside risk over a certain period of time. It is defined as the maximum observed loss in the value of an investment expressed in percentage terms, and it is calculated as the difference between the value of the lowest trough and the value of the highest peak before the trough (graphically in figure 4.5):

$$MDD = \frac{\text{Trough Value} - \text{Peak Value}}{\text{Peak Value}} \tag{4.1}$$

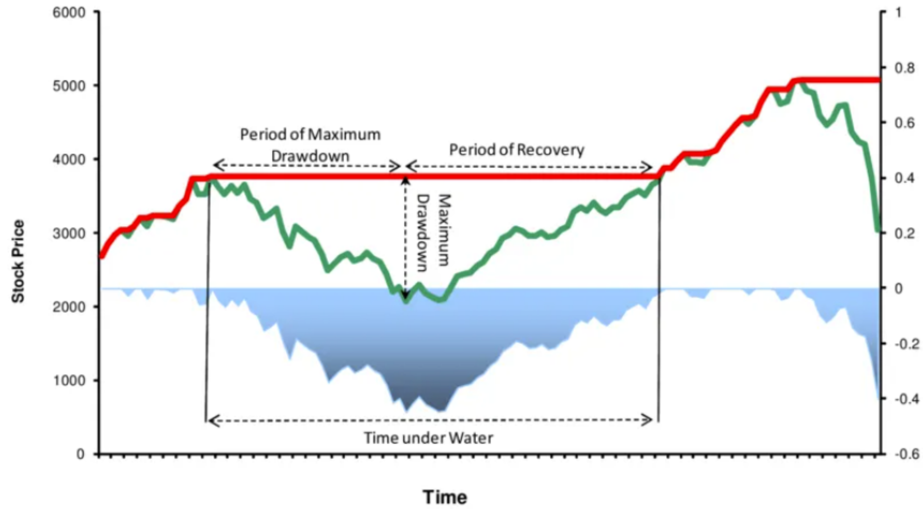
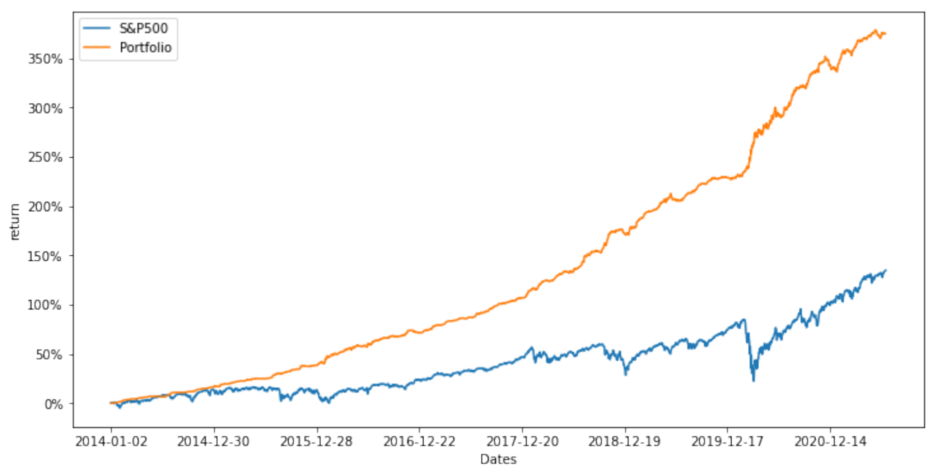


Figure 4.5: Graphical representation of maximum drawdown. Source: Fischer and Lind-Braucher (2009)

4.2.1 U.S. market



Annualized Return	Annualized Vol	Skewness	Kurtosis	Max Drawdown
0.230406	0.04585	1.28422	10.502335	-0.037812

Figure 4.6: Return of the portfolio vs. S&P500 return (above) and portfolio statistics (below) for the entire trading period

Annualized Return	Annualized Vol	Skewness	Kurtosis	Max Drawdown
0.120601	0.175491	-0.678019	23.099533	-0.339251

Figure 4.7: Statistics of the market portfolio over the entire trading period

Returns

Over the entire eight-year period of trading, our portfolio yields a total net return equal to 399.98%, against a return on S&P500 of 134.58%, outperforming it by 265.40%. Focusing on annualized returns to better grasp the magnitude of these numbers, as shown in figures 4.6 and 4.7, our total return would result in an annualized mean return equal to 23.04%, while the one of the market portfolio would result in a mean annual return of 12.06%, leading to an annual mean overperformance of 10.98% in favor of our strategy.

Return distribution analysis

Moreover, looking at the distributions of said returns for the two portfolios, as indicated in figures 4.6 and 4.7, it appears to be clear that they are very far from a normal distribution both in terms of kurtosis and skewness. However, comparing the values of the kurtosis, our portfolio return distribution results to be much less subjected to the presence of extreme values than the market's one, since its kurtosis is much smaller than the other. In particular, even though both distributions are leptokurtic, ours has a kurtosis value equal to 10.5, while the one of market portfolio return is equal to 23.1. Thus, as mentioned when introducing kurtosis, market portfolio distribution will have much fatter tails than our portfolio distribution, making the overall profit more affected by extreme fluctuation in returns.

With regard to the skewness, while market distribution presents a skewness value close to zero (and thus close to the value of a normal distribution) and it is exposed to the downside and upside risk in the same way, our distribution presents a positive skewness equal to 1.28, resulting in an asymmetry to the left side of the median value. Hence, even if the observed values are on average lower than the mean (due to the mode of returns being lower than the mean in distributions with positive skewness), thanks to the length of the right tail, we are less affected by downside risk and more by the upside. In conclusion, associating this kind of skewness with our kurtosis value will result in a right tail of the distribution containing several extreme values and, thereby, capable of positively influencing our overall return.

Volatility

Turning our gaze at the strategy's actual risk, expressed in terms of both volatility and maximum drawdown, previous results (in terms of exposure to downside and upside risks) seem to be supported by values observed in these measures. Indeed, it is interesting to note how, despite granting a much higher return, our portfolio still results to be less risky and more stable than the market. The annual mean volatility (expressed by the standard deviation) is very low and equal to 4.59%, with a maximum drawdown of just 3.78%. On the contrary, the market portfolio attests itself to an annual mean standard deviation of almost 20% (17.55%), with a maximum drawdown ten times higher than the one observed in our portfolio (33.92%). These results definitely confirm what was assessed before when studying kurtosis and skewness values: since our portfolio return distribution has a fat right tail, it results to be exposed to the upside risk and not to the downside, to a certain extent.

Market neutrality

Talking about risk, we want to test now whether or not our portfolio is actually market neutral. To do so, we want to run a linear regression of our portfolio's excess return (or risk premium, r_{RP}) over the market's excess return. Excess returns are defined as the difference between a given rate of return and the related risk-free rate:

$$r_{RP} = r - r_f \tag{4.2}$$

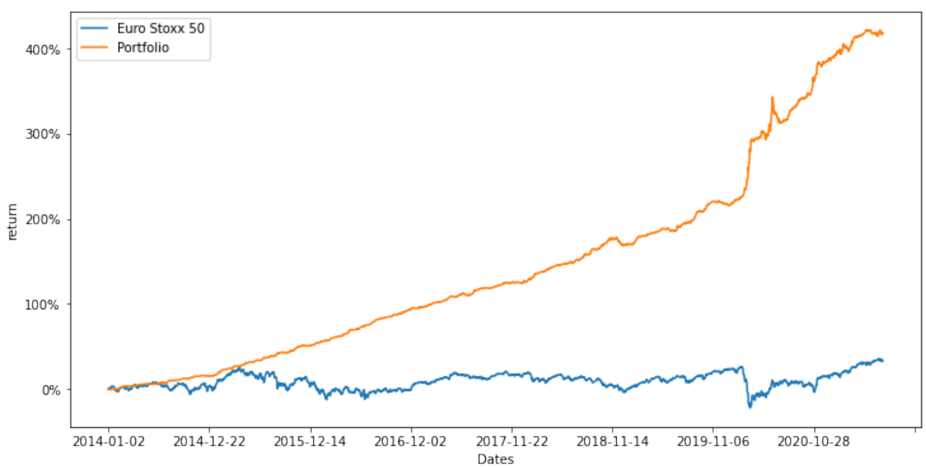
where r_{RP} is the excess return, r is the considered rate of return and r_f is the risk-free rate. In this study, as the risk-free rate for the U.S. market, we want to use the one-year Treasury Bill rate since the chances of the U.S. government going bankrupt within a year are very low. Running the regression will output the following results: It is clear, looking at the regression outputs in table 4.1, that no correlation with the market return is found. In fact, the coefficient beta results to be non-significant since its p-value is equal to 8%

Beta	0.0947
Standard error	0.054
p-value	0.083

Table 4.1: Regression outputs for market neutrality, U.S. market

and so not even acceptable for a confidence interval of 95%. For this reason, we can assess that our portfolio is actually market-neutral, and therefore it is not influenced by market movements.

4.2.2 European market



Annualized Return	Annualized Vol	Skewness	Kurtosis	Max Drawdown
0.236772	0.053486	2.824871	32.092348	-0.070413

Figure 4.8: Return of the portfolio vs. Euro Stoxx 50 return (above) and portfolio statistics (below) for the entire trading period

Annualized Return	Annualized Vol	Skewness	Kurtosis	Max Drawdown
0.037918	0.194785	-0.798136	14.114084	-0.38274

Figure 4.9: Statistics of the market portfolio over the entire trading period

Returns

As it can be easily noted from the graphic in figure 4.8 and as well as in the previous case, our portfolio vastly outperformed its relative market index. It attested on a 420.41% return performance over the whole trading period, versus a total net return of the Euro Stoxx 50 equal to 32.82%, thus generating a vast total net excess return over the market of 387.59%. Moving on to annualized mean net returns for clarity's sake, our portfolio, over the eight-year period, returned on average a 23.68%, while yearly the Euro Stoxx 50 performed, still on average, almost exactly the 20% (19.89%) less attesting on an annual mean net return of 3.79%.

Return distribution analysis

Taking now into account the measures of the distributions of returns indicated in the figures 4.8 and 4.9, also in this case (and, maybe, even more this time), it appears very safe and sound to claim that the two distributions are far away from following a normal one. Comparing values of the kurtosis (14.11 for the market portfolio and a really vast 32.09 for our statistical arbitrage portfolio) reveals that, unlike in the U.S. market, even though the two distributions are still leptokurtic, our portfolio presents fatter tails than the market portfolio, resulting to be more exposed to the presence of extreme values in the distribution of returns.

Considering the skewness value, coming to the skewness, this measure suggests that results found looking at the kurtosis may not be harmful for our strategy. While the market portfolio presents an almost-normal skewness value equal to -0.80, our portfolio has a strongly positive skewness with a value equal to 2.83. Since values of the skewness greater than three are considered symptomatic of a highly asymmetric distribution, we can claim with a certain extent of confidence that our distribution is strongly asymmetric. Hence, it will present a substantial concentration of the observation in the left side of the distribution, thus generating a long right tail. As seen before in the U.S. market, a high value in kurtosis along with a high positive value of skewness will make our portfolio less likely to be exposed to the downside risk and more likely to benefit from the upside one.

Volatility

Looking at the volatility, expressed in terms of annualized standard deviation and also taking into account the maximum drawdown, once again, previous results seem to be confirmed by the values of these volatility measures. Thus, while our portfolio is more rewarding in terms of return, it is also less risky than the market portfolio since its standard deviation is an annual 5.35% against an annual standard deviation of the Euro Stoxx 50 portfolio of 19.48%. Moreover, our portfolio is less risky also in terms of smaller losses: in fact, its maximum drawdown is 7.04%, against a maximum drawdown scored by the market portfolio of 38.27%. Such results tend to confirm and to be in line with what we previously stated: despite the very high skewness and kurtosis that would lead to the presence of extreme values in the distribution, it seems that our portfolio while benefiting from the upside risk, is less exposed to downside risk.

Market neutrality

Having highlighted the main statistics of the portfolio, we now want to check whether it is market neutral or not, in order to try to explain part of the return with a possible correlation with the market. To do so, we have to run the regression of the excess return of our portfolio against the excess return of the market portfolio. In this work, as the European market lacks of a unanimously acknowledged risk-free rate, we considered the one-year rate on the German Bund as a reliable proxy of it since Germany represents one of the most solid economies in the European Union.

Beta	0.0818
Standard error	0.052
p-value	0.121

Table 4.2: Regression outputs for market neutrality, European market

From the outputs of the regression reported in table 4.2, we can state that, also in this case, our strategy is market neutral: the p-value is equal to 12.1%, so the beta is not significant for any reasonable confidence interval. On top of this, we cannot explain any part of our portfolio return with a correlation to the market and, in particular, with the risk premium associated with bearing the market risk.

4.2.3 Sharpe ratios

After having introduced the "raw" performance of the strategy in the two different markets, now we may want to use a risk-adjusted measure of performance in order to make the two analyses more comparable one to another. In this study, such a measure of performance is the Sharpe ratio.

Sharpe ratio (SR) was developed by William Sharpe and is used in order to evaluate the return of an investment, or a portfolio, in accordance with its risk. It represents the average earned excess return over the risk-free rate per unit of risk:

$$SR = \frac{r_p - r_f}{\sigma_p} \tag{4.3}$$

Where r_p is the portfolio return, r_f is the risk-free rate, and σ_p is the standard deviation of the portfolio.

Values of the ratio above one are considered good, and values over three are considered excellent. Expressed in this way, the Sharpe ratio allows for ranking different assets and portfolios according to their Sharpe value. Typically, nevertheless, only peers are compared using the ranking system.

Applying this ratio to the performance of our two statistical arbitrage portfolios, we will have:

- for the U.S market:

$$SR_{us} = \frac{r_{p,us} - r_{f,us}}{\sigma_{p,us}} = 5.00 \tag{4.4}$$

- for the European market:

$$SR_{eu} = \frac{r_{p,eu} - r_{f,eu}}{\sigma_{p,eu}} = 4.59 \tag{4.5}$$

Where $r_{p,us}$ and $r_{p,eu}$ are the returns of our statistical arbitrage portfolios in the U.S. and European market, respectively, $r_{f,us}$ and $r_{f,eu}$ are the risk-free rate for the U.S. and the European market, and $\sigma_{p_{,us}}$ and $\sigma_{p_{,eu}}$ are the standard deviation of the portfolio still in the two different markets.

As we see from equations 4.4 and 4.5, Sharpe ratios of both portfolios are attesting over a value of four, thus classifying them as excellent. Comparing the two Sharpe ratios, the portfolio structured by trading on the U.S. market appears to be better than the other. Indeed, from the comparison between the two performances shown in figures 4.6 and 4.8, it can be noticed that the portfolio trading in the U.S. market performs only slightly worse in terms of return but way better in terms of risk, making it a better option.

In order to try to explain these excellent performances, in the next chapter, we will analyze the risk exposure of our strategy, trying to assess the causes of these returns.

Chapter 5

Risk exposure

In this final chapter, we are going to analyze our returns in order to explain them: firstly, we will go through all the risks to which our strategy may be exposed, defining each of them thoroughly. Then, coming to the market liquidity risk, which configures itself as the main risk for our portfolio, we are going to briefly dwell on the Fama-French three-factor model to introduce the model made by Pastor and Stambaugh with which we will analyze our abnormal returns, trying to assess how much of them are explainable with the exposure to said risk.

5.1 Key risks

Despite the fact that our portfolio is basically market neutral, this does not mean that it is not exposed to any kind of risk. Indeed, oftentimes, very high returns are explainable with the positive relationship that bounds return to risk.

Since statistical arbitrage belongs to the family of long/short equity strategies, it will be exposed to all their typical risks and also to some risks inherent to this specific strategy. In particular, we have:

- specific risk
- market risk (whenever we are not perfectly market-neutral)
- recall risk
- margin risk

- market liquidity risk

5.1.1 Specific risk and Market risk

Specific risk

Specific risk is the intrinsic risk of every stock or security in general. According to the CAPM, the risk of every stock is expressed by its own beta, which measures the expected change in the stock's return for every one percentage point change in the market return. Thus, the expected return of a security varies linearly with the beta of that security. Stocks with $\beta > 1$ will be riskier than the market, and the contrary is true for stocks with a $\beta < 1$.

However, still according to the CAPM, a stock's beta expresses only the relevant risk of the stock (used to determine the expected return while in equilibrium) and not its risk in its entirety. In fact, in equilibrium, the risk of a stock i can be calculated as:

$$\sigma_i^2 = \beta_i^2 \sigma_M^2 + \sigma_{\epsilon_i}^2 \tag{5.1}$$

Where σ_M^2 is the variance of the market, and $\sigma_{\epsilon_i}^2$ is the variance of the error term obtained from the computation of the expected return.

As can be noticed, the equation 5.1 has two components:

1. $\beta_i^2 \sigma_M^2$
2. $\sigma_{\epsilon_i}^2$

The former represents the market risk, which can be nullified through the use of market-neutral strategies. The latter represents the specific risk (or idiosyncratic risk) of that stock, which will remain on the investor unless he well-diversifies his portfolio.

In this work, we will assume our portfolio is well-diversified, as it includes 40 stocks and after 30, benefits from diversification tend to be marginal.

Market risk

If components of traded pairs do not belong to the same or similar sectors, and thus their betas are sensibly different, the beta of our portfolio (β_p) will automatically be not exactly zero or, at least, in the range of it, and therefore negligible. Whenever this happens, our portfolio is no longer market-neutral, and its return will be correlated with the market in a certain measure, which means that the portfolio is exposed to the market risk.

5.1.2 Recall risk

Recall risk is a typical risk of short-selling. While short-selling a security, a trader would need to borrow that security from a security lender, sell it on the market, and lastly, when its price is lower than the price at which he sold it, rebuy the security on the market in order to close the position and return the stock to the security lender.

His profit will then be given by the difference:

$$r = p_1 - p_2 - c \tag{5.2}$$

where p_1 is the price at which the security is sold to a third party, p_2 is the price at which the trader rebuys it on the market to close his position, and c is the lending fee applied by the security lender.

The recall risk refers to the risk that the security lender could recall the borrowed stock at any moment, forcing the trader to close his position earlier and possibly incur losses.

Let us make an example:

A trader wants to short a given stock ABC, whose price is 100\$. In order to do so, he borrows that security for one month with a flat lending fee of 2\$. After one month, the price of stock ABC is 97\$, the trader closes his position and secures a profit of 1\$. Since, according to equation 5.2:

sells at p_1	100\$ -
buys at p_2	97\$ -
pays c	2\$ =
profit	1\$

However, if the security lender decided to recall the stock after only two weeks when the price of A was 99$, the trader would have been forced to close the position realizing a loss of 1$.

sells at p_1	100$ -
buys at p_2	99$ -
pays c	2$ =
profit	-1$

5.1.3 Margin risk

To explain this kind of risk, firstly, we need to introduce the concept of margin.

Margin

In financial transactions, the margin is the collateral that an investor has to deposit at the broker when opening certain positions. This deposit is made in order to cover the credit risk deriving from adverse fluctuations of the market, which would make the investor incur into a loss. In other words, the margin can also be seen as the amount of own funds the investor uses in a given operation.

Definition of margin risk

Margin risk is the risk that the value of the margin could change over time due to fluctuations in the value of the activities deposited as collateral. The value of the margin is calculated daily, following a marking-to-market logic similar to the one explained when calculating returns in paragraph 3.5. If its value falls under a certain threshold set by the broker, or it is no longer sufficient to cover any possible losing position, a margin call will occur.

A margin call is the request from the broker to the investor of depositing additional funds, in terms of money or securities, in order to bring the margin back to its minimum required amount, which is known as maintenance margin. Let us make an example.

Suppose an investor buys $10000 of Alphabet Inc. using $5000 of his own funds and borrowing the remaining \$5000 from the broker, which in turn, requires a maintenance margin of 25%. Since the amount of investor's percentage of own funds (OFP) is calculated as:

$$\text{OFP} = \frac{\text{OF}}{\text{MVS}} = \frac{\text{MVS} - \text{BF}}{MVS} = \frac{10000 - 5000}{10000} = \frac{1}{2} = 50 \tag{5.3}$$

Where OF, MVS and BF are the investor's own funds, the market value of securities, and the amount of borrowed funds, respectively. Since 50% is far above the maintenance margin of 25%, the investor doesn't need to do anything else when opening his position. Suppose that two weeks later, the value of purchased securities falls to \$6000, then the value of investor's equity will shrink to \$1000, resulting in a margin value of 16.67% of the position:

$$\text{OFP} = \frac{6000 - 5000}{6000} = \frac{1}{6} = 16.67\% \tag{5.4}$$

Now, this is below the maintenance margin of 25%. Therefore, the broker makes a margin call and requires the investor to deposit additional funds to restore the minimum required margin level (25%). In particular, the investor will have to deposit \$500, since:

$$\text{AD} = (\text{MVS} \cdot \text{MM}) - \text{OF} = (6000 \cdot 25\%) - 1000 = 500 \tag{5.5}$$

Where AD is the additional deposit required by the broker, and MM is the maintenance margin.

At this point, the investor can either decide to restore the maintenance margin or close part of his open positions in order to match the said margin on the new total value of the position.

5.1.4 Market liquidity risk

Market liquidity risk manifests as the inability to buy or sell a certain amount of a given stock at the desired price. Obviously, this is a more common phenomenon when dealing with large positions: if we own a real estate \$10000000 worth, and there are no "big buyers" on the market since it is going through a recession, if we wanted to sell the asset immediately we would have to sell it at a price smaller than its value.

Even though market liquidity risk may be more obvious when dealing with significant positions, it also exists for small ones depending on the traded asset class and the market's depth.

Market liquidity risk is one of the main risks for statistical arbitrage strategies since they rely on opening and closing positions at very precise price levels and often trade very large volumes.

Due to its importance, and to the fact that our portfolio is not actually exposed to previous risks[1], this risk and its impact on the performance of the strategy will be analyzed more deeply in the following section.

5.2 Exposure to market liquidity risk

As mentioned before, market liquidity risk is the risk of being unable to sell or buy the desired amount of securities at market price. This risk can be influenced by factors like:

- Asset type: Very intuitively, simple assets are more liquid than complex assets. Thus, for instance, a simple stock will typically be more liquid than a CDS[2].
- Substitution: If an asset can be substituted with another instrument, and the substitution cost is low, then the liquidity tends to be higher for that asset.
- Market microstructure: Stocks and derivatives regulated markets are typically much more liquid and deep than many over-the-counter (OTC) markets, which can to be very thin [3].
- Time-frame: If the seller can wait to sell his assets, then market liquidity risk does not represent a threat. On the contrary, if the seller has an urgency to sell, market liquidity risk can be a problem.

The liquidity of an asset or a market can be measured in different ways. For example, while being a simple measure, the bid-ask spread can be indicative of the market depth: more liquid assets or markets will have narrower bid-ask spreads. Or, again, the sensitivity of an asset price to the presence of big long or short orders: the more it is sensitive to them, the more it is illiquid.

[1] Specific risk exposure is eliminated through diversification; market risk is not a problem with market-neutral strategies since the portfolio β is zero; and lastly, in this study, recall risk and margin risk are neglected for simplicity purposes.

[2] i.e., Credit Default Swap

[3] A market is said to be thin when it is characterized by a low number of buyers and sellers for a certain period of time.

Moreover, since the inverse relationship that bounds risk and return, whenever an investor bear this kind of risk, he should receive a risk premium for this.

In this regard, as we will discuss in the next paragraph, Lubor Pastor and Robert Stambaugh, in their paper "Liquidity risk and expected stock returns" (2003), came out with a model which, through the use of a certain measure of liquidity, analyzed stock returns assessing the portion that could be attributed to the risk premium earned for bearing the market liquidity risk.

5.2.1 Pastor and Stambaugh's model

As we said above, according to Pástor and Stambaugh (2003), expected returns should take into account the risk premium derived from bearing the market liquidity risk. In order to analyze this phenomenon, the two researchers developed a four-factor model, adding a liquidity factor to the famous three-factor model made by Fama and French in 1992, with the aim of explaining changes in stock returns corresponding to changes in aggregate liquidity of the market, and found that expected stock returns are related cross-sectionally to the sensitivities of returns to fluctuations in aggregate liquidity. Their idea is that if stock price movements tend to be partially reversed on the following day, we can conclude that part of the original price change was not due to changes in the stock's intrinsic value, but a symptom of price impact associated with the original trade. Reversals suggest that part of the original price change was a concession by traders, who needed to offer higher purchase prices or accept lower selling prices to complete their trades according to a time constraint.[4]

Said analysis can be performed simply through the use of a multivariate linear regression:

$$r_{i,t} = \alpha_i^0 + \beta_i^M MKT_t + \beta_i^S SMB_t + \beta_i^H HML_t + \beta_i^L L_t + \epsilon_{i,t} \tag{5.6}$$

Where r(i,t) is the excess return of our portfolio over the risk-free rate, the first three factors are the ones from the Fama-French model, and L is the liquidity factor. In particular:

[4] Bodie et al., 2009, chap. 13

1. MKT: this is the market factor and denotes the excess return of a broad market index over the risk-free rate (SP500 and Euro Stoxx 50 in this work)

2. SMB: the "small minus big" factor is represented by the payoff, on long-short spreads constructed by sorting stocks according to their market capitalization, that smaller market capitalization companies have over larger companies in terms of excess return. This is also known as the "size effect".

3. HML: this is the "high minus low" factor, and as for SMB, it is represented by the payoff on long-short spreads, but this time constructed by sorting stocks according to their book-to-market ratio, which compares a company's book value to its corresponding market value. The book value is the value of assets netted by the value of the liabilities, and the market value is the market price of the company's shares multiplied by the total number of shares outstanding. This factor is also known as the value factor since stocks with high book-to-market ratios are called "value stocks".

4. L: is the liquidity factor, added to the model by Pastor and Stambaugh, which represents the payoff on the long-short spreads of a portfolio constructed by ranking stock according to the sensitivity of their prices to changes in aggregate liquidity of the market (cfr. figure 5.1).

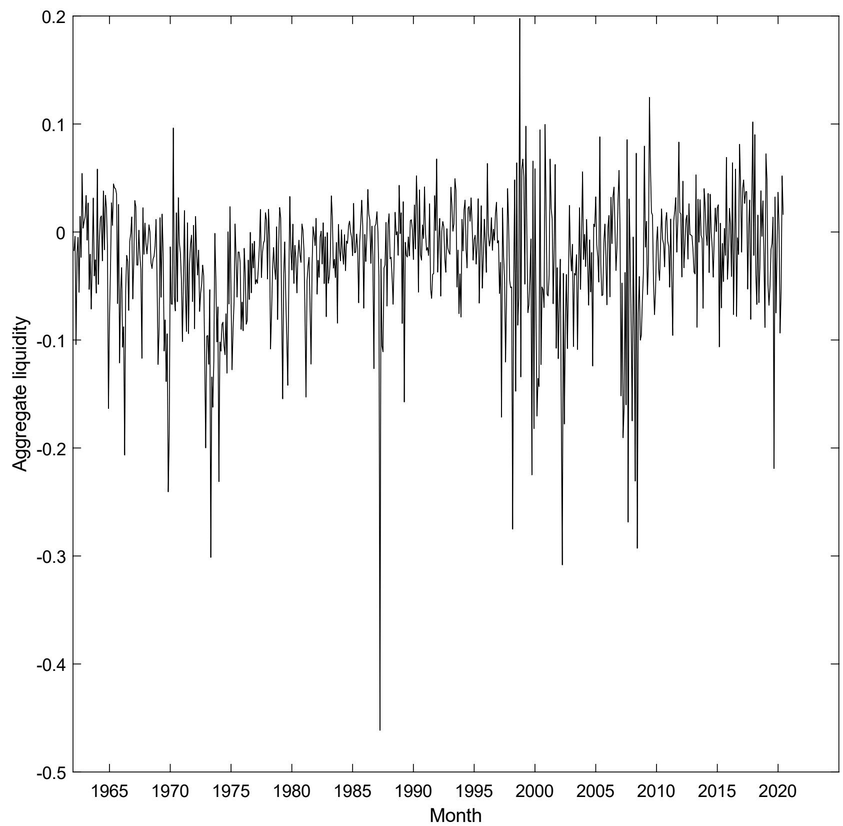


Figure 5.1: Aggregate liquidity series (source: Robert Stambaugh)

After having introduced the model, we will now go through the analysis running the test on our sample, dividing it according to the specific market: U.S market and European market.

Note that all data for Fama-French factors and the liquidity factor are monthly data, and were downloaded from "Kenneth French - Data Library" and the official page of Robert Stambaugh on the University of Pennsylvania's official web domain, respectively. Unfortunately, liquidity factor data for 2021 were unavailable. Hence the analysis will stop at the end of December 2020.

U.S. market

After calculating monthly returns for both the S&P500 (r_M^{sp500}) and our portfolio (r_p^{sp500}), we can use the one-year Treasury Bill rate as the risk-free rate (r_F^{us}) for a U.S.-based investor in order to calculate the excess returns of said series, and put these terms in the regression 5.6:

$$(r_{p,t}{}^{us} - r_F^{us}) = \alpha_p^0 + \beta_p^M (r_M^{us} - r_F^{us})_t + \beta_p^S SMB_t + \beta_p^H HML_t + \beta_p^L L_t + \epsilon_{p,t} \quad (5.7)$$

Running the regression we will estimate the parameters as in table 5.1.

	Coeff.	**Standard err.**	**p-value**
MKT	0.0009	0.002	0.658
SMB	0.0018	0.004	0.621
HML	0.0021	0.003	0.440
L	0.3139	0.290	0.282
α	-0.0458	0.008	0.000

Table 5.1: Regression outputs, U.S.

As we can see from table 5.1, even with a confidence interval (c) of 90% confidence interval we pick, all the coefficients of the regression result to be non-significant since all of their p-values are far above the 20%[5]. Thus, the only significant parameter is the alpha[6], which would tell if our strategy is overperforming, if positive, or underperforming, if negative, like in this case, the market. However, since all the regression factors are non-significant, even though the alpha is significant and negative, and it would mean that we are underperforming the market, this parameter alone cannot be taken into account and is not indicative of the performance as confirmed by the comparison between the performance of our portfolio with the one of the market.

The fact that none of these factors are not adequate in explaining our returns does not mean that the test was inconclusive; on the contrary, the statistical insignificance of the three factors from the Fama-French model

[5] $1 - c = 1 - 0.9 = 0.1 = 10\%$, and $20\% > 10\%$

[6] which would highlight the skill of the asset manager in stock picking and in generating a return that the other factors cannot explain

confirms and is in accordance with the fact that we are following a market-neutral strategy. Furthermore, coming to the liquidity factor, it may not be adequate in explaining our returns since the dimension of the positions opened in this study is moderate. Thus, especially in the U.S. stock market, which is a widely traded, very deep, and structured market, the market liquidity risk tends to become a real problem only for massive orders.

European market

In order to test the exposure to the market liquidity risk on the European market, we need to do all the same steps done for the U.S. one. This time our monthly portfolio return (r_p^{eu}) and market return of the Euro Stoxx 50 (r_M^{eu}) are used to calculate the excess returns, along with the one-year rate on Germany's Bund, taken a the risk-free rate (r_F^{eu}). So the regression will be:

$$(r_{p,t}{}^{eu} - r_F^{eu}) = \alpha_p^0 + \beta_p^M (r_M^{eu} - r_F^{eu})_t + \beta_p^S SMB_t + \beta_p^H HML_t + \beta_p^L L_t + \epsilon_{p,t} \quad (5.8)$$

running it, it will output the parameters as in table 5.2.

	Coeff.	**Standard err.**	**p-value**
MKT	0.1148	0.040	0.006
SMB	$1.98 \cdot 10^{-5}$	0.001	0.984
HML	$6.901 \cdot 10^{-5}$	0.001	0.927
L	0.2612	0.081	0.002
α	0.0168	0.002	0.000

Table 5.2: Regression outputs, Europe

As we can see from the table, also in this case, SMB and HML are not significant for every confidence interval. However, liquidity factor, market factor and the alpha are significant even for a confidence interval (c) of 99%, since their p-values are smaller than 1%. Therefore, this means that these two factors and the intercept (α) are helpful in explaining our returns, and in particular:

- $MKT = 0.1148$ means that investing €100 in our portfolio would be equivalent to invest €11.48 in the market portfolio and €88.52 in the risk-free asset. This factor loading confirms that our strategy in the European market is not perfectly market-neutral; instead, its return presents some sort of correlation with market returns.

- $L = 0.2612$ means that our portfolio will be more tilted towards stocks highly sensitive to aggregate liquidity shocks. Therefore, differently from the situation in the U.S. market, we result to be exposed to the market liquidity risk, this will generate a higher risk premium (due to exposure to said risk) and lead to a higher spread return between our portfolio and the benchmark index as seen in chapter 4.2.3.

- Even though the two previous factors help explain part of the return, they cannot explain it in its entirety. In fact, the intercept alpha tells us that there is a percentage of the return that these factors cannot explain. In particular, $\alpha = 0.0168$ means that we are generating, every month, an additional return of 1.68% due to skills in stock picking and other factors.

Even though our exposure to market liquidity risk is not dramatically high, in the European stock market, we can still notice its effects, and this is probably due to the fact that, compared to the U.S. stock market, its trading volumes are much smaller and the market itself is not as deep and structured as the U.S. one. Other than that, our results confirmed our previous analysis presented in chapter 4.2.3 depicting our portfolio as very solid, especially in terms of risk exposure.

Chapter 6

Conclusions

In the course of this dissertation, we thoroughly described the particular strategy behind our cointegration-based statistical arbitrage portfolio, deeply explaining and defining the statistical concepts lying behind it and eventually managed to show how a statistical arbitrage strategy can be very profitable while actually reducing the volatility of returns and the exposure to the downside risk in general.

Furthermore, as emerged from chapter 4.2.3, despite the portfolio trading in the U.S. market performed better than its European counterpart, it was evident (and widely confirmed by the Sharpe ratios equal to 5.00 and 4.59, respectively) that the strategy was effective and profitable in both markets, guaranteeing stable returns, granted by the market neutrality of the portfolios, even in periods of high volatility or crisis of the reference market. In particular, as shown in section 4.2, over the entire trading period, the portfolio trading in the U.S. market yielded a lower total net return than the one trading in the European market, scoring a return equal to 399.98%, against a solid 420.41%. Nevertheless, about the volatility, the U.S.-based portfolio registered an annual standard deviation of 4.59% with a maximum drawdown equal to just 3.78% against a worse performance of the Europe-based portfolio, which registered an annual standard deviation of 5.35% along with a maximum drawdown equal to 7.04%. Even though the European portfolio performed a bit worse than the other with regard to the overall performance, it is interesting to note how, in this case, the performance of the portfolio results much higher than the one of its reference market: the Europe-based portfolio generated a remarkable total net excess return over the market equal to 387.59%.

Given the presence of these abnormal returns, through the four-factor model developed by Pastor and Stambaugh, we tried to assess the factors influencing them to explain their level, and particularly, we tried to assess whether these returns were incorporating a risk premium because affected by the market liquidity risk. According to the results, presented in tables 5.1 and 5.2, it appears that there is no inefficiency in the U.S. market, and an investor following this strategy should not be exposed to the market liquidity risk. On the other hand, since in the European market the liquidity factor L is significant and equal to 0.2612, our portfolio returns are affected by the liquidity risk, forcing a potential investor following this strategy to bear it. Considering these results, the reason for the abnormal performance becomes clear: the higher return of the European portfolio was due to incorporating a risk premium for bearing the market liquidity risk, and the presence of this risk caused the measures of volatility to be higher than the ones of the U.S.-based portfolio.

The results presented in this study may be considered optimistic since they do not consider any trading costs such as lending fees and transaction costs. However, as presented by Gatev et al. (2006), whenever dealing with highly traded and liquid assets (as in our case, since in the U.S. portfolio we do not suffer the liquidity risk), transaction costs and lending fees tend to assume minimal values, in the range of 30-40 basis points. On top of this, despite considering such costs would inevitably affect returns and the performance in general, looking at the magnitude of these numbers, it is probably safe to assert that the strategy's validity would remain intact.

In conclusion, given the excellent results obtained, this statistical arbitrage strategy proved (still with all the assumptions of the case) to be solid and reliable even throughout periods of market turbulence. The algorithmic tool proved to work and to be efficient in backtesting the strategy. It will be tested in reality, optimized, and improved in the future. Indeed, there is room for improvements and additional developments as, for instance, adding I.A.-based systems to rank pairs, using machine learning and deep learning to set more precise entry levels for the positions. As a final consideration, according to what we mentioned in chapter 2, we would state that a more thorough test of cointegration, like Johansen's one that relies on trading sets of stocks instead of pairs, would be worth to be explored and could represent a more powerful and sophisticated evolution of the strategy to guarantee more robust returns with lesser risk exposition.

Appendices

.1 Codes

This section will report some of the functions used for the carried-out analysis, briefly commenting on them. For the sake of brevity, we will not report the entire code of the algorithm since they are more than one thousand lines of code. For all of the following codes the numpy package is imported as np, and the pandas package as pd.

.1.1 Testing for stationarity

```
    def find_stationary(data, cutoff=0.05):

        stationary = []
        for i in range(0,len(data.columns)):
        pvalue = adfuller(data.iloc[:,i])[1]
        if pvalue <cutoff:
        stationary.append(data.columns[i])
    return stationary
```

This function checks if a series is stationary, and if yes, adds it to a list to be dropped from the dataset.

.1.2 Finding cointegrated pairs

```
  def find_cointegrated_pairs(data,sortby = "half_life"):
    n = data.shape[1]
    score_matrix = np.zeros((n,n))
    pvalue_matrix = np.ones((n,n))
    keys = data.keys()
    pairs = []
    score_pairs=[]
    pval_pairs = []
    for i in range(n):
        for j in range(i+1,n):
            Series1 = data[keys[i]]
            Series2 = data[keys[j]]
            result = coint(Series1, Series2)
            score = result[0]
            pvalue = result[1]
            score_matrix[i,j] = score
            pvalue_matrix[i,j] = pvalue
            if pvalue < 0.05:
```

```
                pairs.append((keys[i],keys[j]))
                score_pairs.append(score)
                pval_pairs.append(pvalue)

    hfs,pva = halflife_pairs(data,pairs)

    s = np.array(score_pairs)
    s_n = (s-np.mean(s))/np.std(s)
    h = np.array(hfs)
    h_n = (h-np.mean(h))/np.std(h)
    skore = s_n + h_n

    ranked_pairs = pd.DataFrame(data={"pair":pairs,"score":
    skore,"half_life":hfs,"score_test":score_pairs,
    "significance":pva}).sort_values(by=sortby)
    drop_list = ranked_pairs.loc[ranked_pairs.significance
   >0.05].index.to_list()
    ranked_pairs.drop(index=drop_list,inplace=True)
    ranked_pairs.reset_index(drop=True,inplace=True)
    return score_matrix, pvalue_matrix, pairs, ranked_pairs
```

This function test all the pairs in the dataset with each other in order to assess if there is cointegration between them. Once found all the possible cointegrated pairs, they are ranked according to their mean reversion capability.

.1.3 Generation of the beta for cointegrated pairs to open the positions

```
    def generate_betas(y_s,x_s,df_in):
    betas = []
    for i in range(len(y_s)):

        y = df_in.loc[:,y_s[i]]
        x = df_in.loc[:,x_s[i]]
        x = sm.add_constant(x)
        model= sm.OLS(y,x)
        results = model.fit()
        beta = results.params[1]
        betas.append(beta)
    return betas
```

With this functions, we calculate betas for every pair in order to have the dimension of the different long and short positions.

.1.4 Signal matrix

```
def signal_matrix(data, spread_matrix,n_std=1):

n = spread_matrix.shape[0]
m = spread_matrix.shape[1]
signal = np.zeros((n,m))
n_transaction = []




for i in range(0,len(spread_matrix.columns)):

    t_up = 1
    t_down = -1

    signal_array = np.zeros(len(signal))


    for j in range(0,len(signal)):

        thresh_close = 0

        sp = spread_matrix.reset_index(drop=True)
        val = sp.iloc[j,i]


        no_positions = np.nansum(signal_array[:j])==0
        long_positions = np.nansum(signal_array[:j])==1
        short_positions = np.nansum(signal_array[:j])==-1

        #OPEN LONG
        if val < t_down and no_positions:
            signal_array[j] = 1

        #OPEN SHORT
        if val > t_up and no_positions:
            signal_array[j] = -1


        #CLOSE LONG
        if val > thresh_close and long_positions:
            signal_array[j] = -1

```

```python

            #CLOSE SHORT
            if val < thresh_close and short_positions:
                signal_array[j] = 1

            if j == len(signal)-1:
                if long_positions:
                    signal_array[j] = -1
                if short_positions:
                    signal_array[j] = +1

        signal[:,i] = signal_array.T
        s = pd.DataFrame(data=signal)
        s.columns = spread_matrix.columns
        #
        s.sum()
    return s
```

This is the function that creates our signal matrix. Starting from a null-matrix it substitues zeroes with $\pm ones$ whenever a given position has to be opened or closed.

.1.5 Maximum drawdown

```python
    def drawdown(return_series: pd.Series):
    """Takes a time series of asset returns.
       returns a DataFrame with columns for
       the wealth index,
       the previous peaks, and
       the percentage drawdown
    """
        wealth_index = 1000*(1+return_series).cumprod()
        previous_peaks = wealth_index.cummax()
        drawdowns = (wealth_index - previous_peaks)/
  previous_peaks
        return pd.DataFrame({"Wealth": wealth_index,
                             "Previous Peak": previous_peaks,
                             "Drawdown": drawdowns})
```

This function is used to calculate the maximum drawdowns.

.1.6 Sharpe ratio

```python
    def sharpe_ratio(r, riskfree_rate, periods_per_year):
     """
     Computes the annualized sharpe ratio of a set of returns
     """
         # convert the annual riskfree rate to per period
         rf_per_period = (1+riskfree_rate)**(1/periods_per_year)-1
         excess_ret = r - rf_per_period
         ann_ex_ret = annualize_rets(excess_ret, periods_per_year)
         ann_vol = annualize_vol(r, periods_per_year)
     return ann_ex_ret/ann_vol
```

This is used to calculate the Sharpe ratios.

.1.7 Skewness and kurtosis

```python
def skewness(r):
    """
    Alternative to scipy.stats.skew()
    Computes the skewness of the supplied Series or DataFrame
    Returns a float or a Series
    """
    demeaned_r = r - r.mean()
    # use the population standard deviation, so set dof=0
    sigma_r = r.std(ddof=0)
    exp = (demeaned_r**3).mean()
    return exp/sigma_r**3
```

```python
def kurtosis(r):
    """
    Alternative to scipy.stats.kurtosis()
    Computes the kurtosis of the supplied Series or DataFrame
    Returns a float or a Series
    """
    demeaned_r = r - r.mean()
    # use the population standard deviation, so set dof=0
    sigma_r = r.std(ddof=0)
    exp = (demeaned_r**4).mean()
    return exp/sigma_r**4
```

These two function are used to calculate the skewness and the kurtosis, respectively.

Bibliography

Y. Amihud. Illiquidity and stock returns: cross-section and time-series effects. *Journal of financial markets*, 5(1):31–56, 2002.

Z. Bodie et al. *Investments*. Tata McGraw-Hill Education, 2009.

R. F. Engle and C. W. Granger. Co-integration and error correction: representation, estimation, and testing. *Econometrica: journal of the Econometric Society*, pages 251–276, 1987.

E. F. Fama and K. R. French. *Common risk factors in the returns on stocks and bonds*. University of Chicago Press, 2021a.

E. F. Fama and K. R. French. *Multifactor explanations of asset pricing anomalies*. University of Chicago Press, 2021b.

E. Fischer and S. Lind-Braucher. Optimal portfolios with traditional and alternative investments: An empirical investigation. *The Journal of Alternative Investments*, 13, 12 2009. doi: 10.3905/jai.2010.13.2.058.

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

R. Hyndman and G. Athanasopoulos. *Forecasting: Principles and Practice*. OTexts, Australia, 2nd edition, 2018.

F.-S. Lhabitant. *Handbook of hedge funds*, volume 332. John Wiley & Sons, 2007.

L. Pástor and R. F. Stambaugh. Liquidity risk and expected stock returns. *Journal of Political economy*, 111(3):642–685, 2003.

G. Vidyamurthy. *Pairs Trading: quantitative methods and analysis*, volume 217. John Wiley & Sons, 2004.

Department of Economics and Finance
Equity markets and alternative investments

Performance Analysis of a Cointegration-Based Statistical Arbitrage Portfolio

Advisor
prof. Paolo Vitale

Co-advisor
prof. Marco Morelli

721221
Federico Cappuccio

Academic Year: 2020/2021

Contents

Chapter 1

Introduction

Algorithmic trading is a rapidly expanding reality. Although it was mainly practiced by banks in the past, more and more retail investors became attracted to it during the years. This phenomenon is, for sure, to be addressed primarily to innovations in the tech industry, along with progress in the A.I. (i.e., artificial intelligence) field, and to the increasing ease of information retrieval for this kind of matter.

The main objectives of this study are to define the characteristics of a statistical arbitrage strategy and analyze the performance of the associated portfolio through the use of a self-developed algorithm, written with Python, which, relying on sophisticated statistical tools, permits to analyze and evaluate blocks of market data for a broad set of activities. Furthermore, the analysis will be carried on in two different markets in order to highlight the change in the performance according to the different structure and characteristics of the reference market.

After introducing the topic and presenting the main features of the strategy, we will then try to investigate the factors that affect the performance.

Chapter 2

Pair trading and statistical arbitrage

2.1 Statistical arbitrage

In developing pair trading's statistical arbitrage, the first was a Wall Street quant in the mid-1980s: Nunzio Tartaglia. Tartaglia, at that time, gathered a team of academics (mainly composed of computer scientists, physics, and mathematicians) in order to develop an automated trading strategy capable of exploiting arbitrage opportunities and market inefficiencies through the systematic use of quantitative and statistical instruments. Tartaglia and his team performed statistical arbitrage for Morgan Stanley in 1987, generating about 50 million dollars of profits for the firm. As time passed, Tartaglia's strategy gathered increasing consensus, becoming today one of the most popular market-neutral strategies among long/short equity funds. [1]

There are several ways to identify tradable pairs in accordance with statistical arbitrage. One of them, and probably among the most famous and simplest quantitative method, is the one utilizing a measure of distance as the tracking variance. However, although this method is straightforward and intuitive, to make more accurate analyses, in this study, we will rely on more complex and thorough measures, such as cointegration.

[1] Gatev et al., 2006

Chapter 3

Trading strategy

In this chapter, we are going to discuss and explain the choices made in structuring the strategy that led us to the construction of our statistical arbitrage-based portfolio. Starting from how a strategy is tested and validated and how data have been chosen, we will gradually pass through all the steps that will lead us to calculate returns and, consequently, measure the portfolio's performance.

3.1 Data and sampling

The study involved daily price data from the first quarter of 2011 (2011Q1) to the second quarter of 2021 (2021Q2) for stock that are part of two major indices: SP500 for the United States market, and Euro Stoxx 50 for the European one. All data were selected from the Refinitiv databases and then sampled into subsets of six months according to the figure 3.1 below, where the in-sample dataset is composed of three years of observations, and the out-of-sample dataset contain six months of them. Every new period each dataset rolls in time over a rolling window of six months in order to allow us to capitalize on information that companies release every semester, helping us build a much more solid and reliable model.

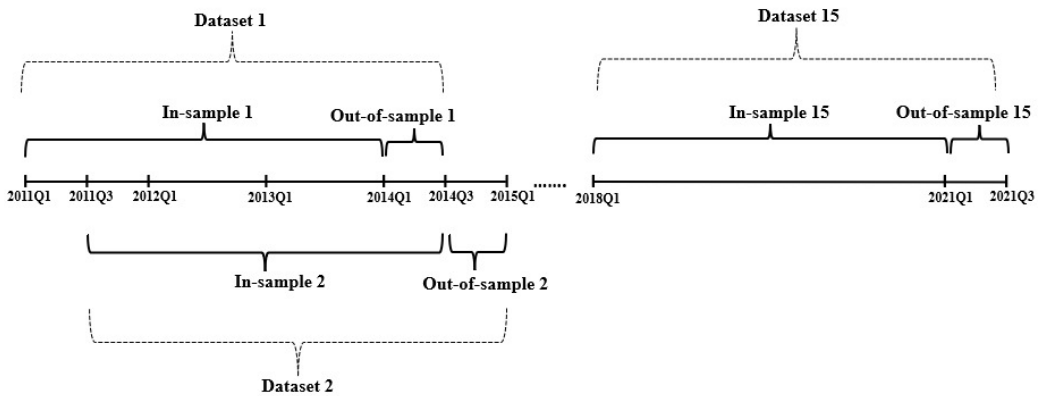


Figure 3.1: Datasets structure

3.2 Data polishing

As we will be performing an Engle-Granger test, we now want to polish datasets from stationary series, dropping them, since such a test cannot be performed on them.

A stationary time series is a series whose properties do not depend on the time at which the series is observed. In general, a stationary time series will have no predictable patterns in the long term. Time plots will show the series as roughly horizontal (although some cyclic behaviour is possible), with constant variance[1].

Testing for stationarity: the Augmented Dickey-Fuller test

In order to detect any stationary series, the augmented Dickey-Fuller test comes in help. This is a statistical significance test that can be used for testing the existence of a unit root in a univariate process in the presence of serial correlation. The null hypothesis of the test asserts that a unit root does exist and, therefore, that the series is not likely to be stationary. The alternative hypothesis is that there is no unit root, and so the series is likely to be stationary. If the p-value is above a critical value, then we cannot reject the null hypothesis, demonstrating the existence of a unit root. Alternatively, the null hypothesis is rejected in favor of the alternative one.

[1]Hyndman and Athanasopoulos, 2018, chap. 8.1

3.3 Pair selection

Once datasets are polished and ready to be used, we want to find a certain number of pairs of stocks, so that each member of a single pair is linked one to the other by a meaningful statistical relationship: the cointegration.

3.3.1 Cointegration

Before cointegration tests were introduced, economists used to rely on linear regressions to find relationships between different time series. In the 1980s, Engle and Granger began to claim that the linear regression approach to the analysis of time series was flawed since it could potentially lead to the so-called spurious correlations[2]. Then in 1987, they coined the term and formalized the concept of cointegration for the first time.

If two time series are integrated of order one and exists a vector beta such that a linear combination of the two time series generates a stationary process, it can be asserted that the two series are cointegrated.

Cointegration test

Many tests can be used in order to assess cointegration between two time series. In this study, we utilize the Engle-Granger test, where the null hypothesis is that there is no cointegration. This method is based on a two-step procedure:

1. performs a static regression of the time series, thus generating the residuals
$$y_t = \mu + \beta_2 s_t + u_t \tag{3.1}$$
where u_t are the residuals of the OLS.

2. tests the stationarity of the residuals checking if a unit root exists by performing an Augmented Dickey-Fuller test.

If a unit root is found in this way and the null hypothesis of the ADF test is accepted,and then there is no cointegration. Contrariwise, if there is no unit root and ADF's H_o is rejected, the two series are cointegrated.

[2]spurious correlations occur whenever two or more variables result to be correlated due to coincidence or unknown factors, conducing to misleading results.

3.3.2 Pair ranking

According to a looping methodology similar to the one utilized for testing stationarity of the series in datasets, the Engle-Granger test is performed on any possible combination of series pairwise. This looping test will lead to a vast number of cointegrated pairs. Thus, in order to maximize the performance, pairs are ranked basing on their mean-reversion capability and then the first twenty are selected.

3.4 Trading signal generation

In order to define the size of long and short positions within each pair, an OLS regression is run, whose beta will tell us how much the values of a given series of stock change when its cointegrated counterpart changes of one unit.

Then for every pair[3], is calculated the spread of the two time series:

$$\Delta(Y_t, X_t) = y_t - x_t \cdot \beta \tag{3.2}$$

and it is standardized to grant that all data have the same scale. ventually, we generate a "spread matrix" containing the standardized time series of the spreads for every pair, along with another matrix (that we will call "signal matrix") in charge of giving the signals of when opening and closing long and short positions on the pairs.

3.5 Portfolio returns

Our returns will be computed by applying multiple weighting schemes, according to the assumptions of the model:

1. Full reinvestment of returns: Gains made both after each trade and at the end of each period are supposed to be totally reinvested, leading to a continuously compounded computation of returns.

2. Equally weighted portfolio at the beginning of every period: At the beginning of every period (i.e., every semester), our statistical arbitrage portfolio will have equal weightings for every couple considered.

[3] each composed by two generic stock Y and X

As time passes and trades are closed, since capital gains are fully reinvested, every single asset its relative weighting according to its own performance. Then, in every new period the portfolio is fully rebalanced to remake it equally weighted.

3. Finite budget: In order to bring the analysis closer to reality, we will assume that our budget is not infinite and represented by a generic amount M, which will be allocated among the pairs in equal parts, corresponding to $M\frac{1}{n}$, where n is the number of considered pairs

4. Perfect divisibility of asset prices: Since assumption three gives us restrictions about our budget, we may not be able to buy the exact quantity that we want of a given stock, hence we want to assume the perfect divisibility of asset prices af if we were trading CFDs[4].

Returns are then calculated in correspondence of opened positions indicated by the signal matrix.

[4]Contracts for differences (CFDs) are contracts between an investor and a financial institution, in which the investor opens a position on the future value of an asset. The difference between the open and closing prices of the trade is cash-settled. CFDs are available for a wide range of assets, such as commodities, shares, and many others

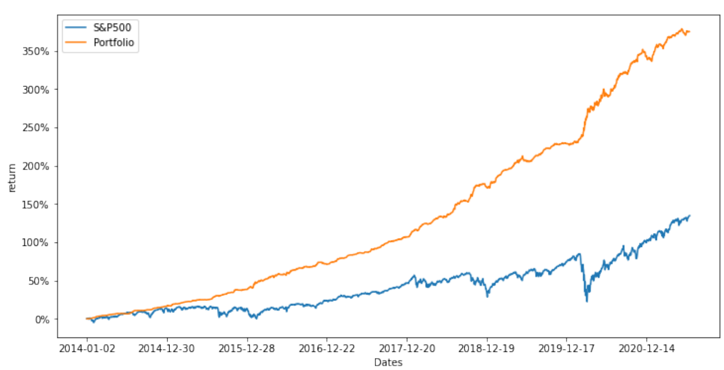
Chapter 4

Performance analysis

In this chapter we will analyze the performance of the strategy.

4.1 Portfolio statistics

4.1.1 U.S. market



Annualized Return	Annualized Vol	Skewness	Kurtosis	Max Drawdown
0.230406	0.04585	1.28422	10.502335	-0.037812

Figure 4.1: Return of the portfolio vs. S&P500 return (above) and portfolio statistics (below) for the entire trading period

Annualized Return	Annualized Vol	Skewness	Kurtosis	Max Drawdown
0.120601	0.175491	-0.678019	23.099533	-0.339251

Figure 4.2: Statistics of the market portfolio over the entire trading period

Returns and their distribution

Over the entire eight-year period of trading, our portfolio yields a total net return equal to 399.98%, against a return on S&P500 of 134.58%, outperforming it by 265.40%. Focusing on annualized returns to better grasp the magnitude of these numbers, as shown in figures 4.1 and 4.2, our total return would result in an annualized mean return equal to 23.04%, while the one of the market portfolio would result in a mean annual return of 12.06%, leading to an annual mean overperformance of 10.98% in favor of our strategy. Moreover, looking at the distributions of said returns for the two portfolios, kurtosis and skewness values are very high in both cases. Due to a kurtosis value of 10.5 and a positive skewness equal to 1.2, our statistical arbitrage portfolio presents a concentration of values on the left side of the distribution with a long and big right tail, resulting thus to be less exposed to the downside risk and more to the upside.

Volatility

Turning our gaze at the strategy's actual risk, expressed in terms of both volatility and maximum drawdown, previous results (in terms of exposure to downside and upside risks) seem to be supported by values observed in these measures. Infact the annual mean volatility (expressed by the standard deviation) of our portfolio is very low and equal to 4.59%, with a maximum drawdown of just 3.78%, against the market portfolio with an annual standard deviation and a maximum drawdown equals to 17.55% and 33.92%, respectively.

Market neutrality

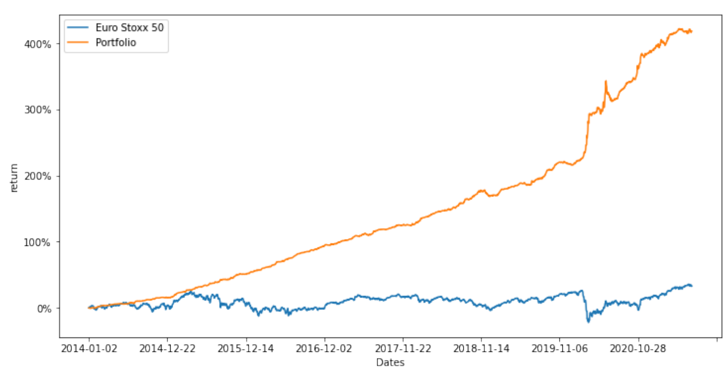
Talking about risk, we want to test now whether or not our portfolio is actually market neutral. To do so, we want to run a linear regression of our portfolio's excess return (over the risk free rate), with the market's excess return. In this study, as the risk-free rate for the U.S. market, we want to use the one-year Treasury Bill rate since the chances of the U.S. government going bankrupt within a year are very low. Running the regression will output the following results:

Beta	0.0947
Standard error	0.054
p-value	0.083

Table 4.1: Regression outputs for market neutrality, U.S. market

It is clear, given the results in table 4.1, that no correlation with the market return is found. In fact, the coefficient beta results to be non-significant, thus we can assess that our portfolio is actually market-neutral.

4.1.2 European market



Annualized Return	Annualized Vol	Skewness	Kurtosis	Max Drawdown
0.236772	0.053486	2.824871	32.092348	-0.070413

Figure 4.3: Return of the portfolio vs. Euro Stoxx 50 return (above) and portfolio statistics (below) for the entire trading period

Annualized Return	Annualized Vol	Skewness	Kurtosis	Max Drawdown
0.037918	0.194785	-0.798136	14.114084	-0.38274

Figure 4.4: Statistics of the market portfolio over the entire trading period

Returns and their distribution

As it can be easily noted from the graphic in figure 4.3 and as well as in the previous case, our portfolio vastly outperformed its relative market index. It attested on a 420.41% return performance over the whole trading period, versus a total net return of the Euro Stoxx 50 equal to 32.82%, thus generating a vast total net excess return over the market of 387.59%. Moving on to annualized mean net returns for clarity's sake, our portfolio, over the eight-year period, returned on average a 23.68%, while yearly the Euro Stoxx 50 performed, still on average, almost exactly the 20% (19.89%) less attesting on an annual mean net return of 3.79%. Taking now into account the measures of the distributions of returns, given the values of kurtosis (32.09) and skewness (2.83), also this time, we have an asymmetric distribution with a high concentration of values in the left side of the distribution and a long and fat right tail. Thus, as asserted for the first case, our portfolio seems to be less exposed to the downside risk and benefit from significant exposure to the upside risk.

Volatility

Looking at the volatility, expressed in terms of annualized standard deviation and also taking into account the maximum drawdown, once again, previous results seem to be confirmed by the values of these volatility measures: even if our portfolio is more rewarding in terms of return, it is also less risky than the market portfolio since its standard deviation is an annual 5.35% with a maximum drawdown of 7.04% against an annual standard deviation of the Euro Stoxx 50 portfolio of 19.48% with a maximum drawdown equal to38.27%.

Market neutrality

Now want to check whether or not our portfolio is market neutral, in order to try to explain part of the return with a possible correlation with the market.

As before, we run the regression of the excess return of our portfolio against the excess return of the market portfolio. In this work, as the European market lacks of a unanimously acknowledged risk-free rate, we considered the one-year rate on the German Bund as a reliable proxy of it since Germany represents one of the most solid economies in the European Union.

Beta	0.0818
Standard error	0.052
p-value	0.121

Table 4.2: Regression outputs for market neutrality, European market

From the results in table 4.2, we can assert that, also in this case, our strategy is market neutral since the beta is not significant for any reasonable confidence interval.

4.1.3 Sharpe ratios

After having introduced the "raw" performance of the strategy in the two different markets, now we may want to use a risk-adjusted measure of performance in order to make the two analyses more comparable one to another. In this study, such a measure of performance is the Sharpe ratio, which is defined as the average earned excess return over the risk-free rate per unit of risk:

$$SR = \frac{r_p - r_f}{\sigma_p} \tag{4.1}$$

Where r_p is the portfolio return, r_f is the risk-free rate, and σ_p is the standard deviation of the portfolio.

Applying this ratio to the performance of our two statistical arbitrage portfolios, we will have a Sharpe ratio equal to 5.00 for the U.S.-based portfolio and a Sharpe ratio equal to 4.59 for the European one.

Comparing the two Sharpe ratios, the portfolio structured by trading on the U.S. market appears to be better than the other, however, as they both score values over four, they are both performing in an excellent way.

Chapter 5

Risk exposure

In this final chapter, we will analyze the exposure to the market liquidity risk, configured as the main risk for our portfolio, through the use of the model developed by Pastor and Stambaugh in 2003, in order to assess how much of our returns are explainable with the exposure to said risk.

5.0.1 Pastor and Stambaugh's model

According to Pastor and Stambaugh (2003), expected returns should take into account the risk premium derived from bearing the market liquidity risk. In order to analyze this phenomenon, the two researchers developed a four-factor model, adding a liquidity factor to the famous three-factor model made by Fama and French in 1992, with the aim of explaining changes in stock returns corresponding to changes in aggregate liquidity of the market, and found that expected stock returns are related cross-sectionally to the sensitivities of returns to fluctuations in aggregate liquidity.[1]

Said analysis can be performed simply through the use of a multivariate linear regression:

$$r_{i,t} = \alpha_i^0 + \beta_i^M MKT_t + \beta_i^S SMB_t + \beta_i^H HML_t + \beta_i^L L_t + \epsilon_{i,t} \tag{5.1}$$

Where r(i,t) is the excess return of our portfolio over the risk-free rate, the first three factors are the ones from the Fama-French model, and L is the liquidity factor.

[1] Bodie et al., 2009, chap. 13

U.S. market

Running the regression for the portfolio trading in the U.S. market, will estimate the parameters as in table 5.1.

	Coeff.	**Standard err.**	**p-value**
MKT	0.0009	0.002	0.658
SMB	0.0018	0.004	0.621
HML	0.0021	0.003	0.440
L	0.3139	0.290	0.282
α	-0.0458	0.008	0.000

Table 5.1: Regression outputs, U.S.

As we can see, the only significant parameter is the alpha[2], however, since all the regressors are non-significant, this parameter alone cannot be taken into account and is not indicative of the performance. So, according to the results, our portfolio is market-neutral and not exposed to the market liquidity risk as well.

European market

In order to test the exposure to the market liquidity risk on the European market, we need to do all the same steps done for the U.S. one. This time our regression will output the parameters as in table 5.2:

	Coeff.	**Standard err.**	**p-value**
MKT	0.1148	0.040	0.006
SMB	$1.98 \cdot 10^{-5}$	0.001	0.984
HML	$6.901 \cdot 10^{-5}$	0.001	0.927
L	0.2612	0.081	0.002
α	0.0168	0.002	0.000

Table 5.2: Regression outputs, Europe

[2] which would highlight the skill of the asset manager in stock picking and in generating a return that the other factors cannot explain

Also in this case, SMB and HML are not significant for every confidence interval. However, liquidity factor, market factor and the alpha are significant.In particular, in the European market, the portfolio results to be exposed to the market risk and the market liquidity risk, justifying the higher levels of risk and return compared to the U.S-based portfolio.

Chapter 6

Conclusions

In the course of this dissertation, we thoroughly described the outline of the particular strategy behind our cointegration-based statistical arbitrage portfolio, deeply explaining and defining the statistical concepts lying behind it and eventually managing to show how a statistical arbitrage strategy can be very profitable while actually reducing the volatility of returns and the exposure to the downside risk in general.

Furthermore, as emerged from chapter 4.1.3, despite the portfolio trading in the U.S. market performed better than its European counterpart, it was evident (and widely confirmed by the Sharpe ratios equal to 5.00 and 4.59, respectively) that the strategy was effective and profitable in both markets, guaranteeing stable returns, granted by the market neutrality of the portfolios, even in periods of high volatility or crisis of the reference market.

According to the results, presented in tables 5.1 and 5.2, it appears that there is no inefficiency in the U.S. market, and an investor following this strategy should not be exposed to the market liquidity risk. On the other hand, since in the European market the liquidity factor L is significant and equal to 0.2612, our portfolio returns are affected by the liquidity risk, forcing a potential investor following this strategy to bear it. Considering these results, the reason for the abnormal performance becomes clear: the higher return of the European portfolio was due to incorporating a risk premium for bearing the market liquidity risk, and the presence of this risk caused the measures of volatility to be higher than the ones of the U.S.-based portfolio.

In conclusion, given the excellent results obtained, this statistical arbitrage strategy proved (still with all the assumptions of the case) to be solid and reliable even throughout periods of market turbulence. The algorithmic tool proved to work and to be efficient in backtesting the strategy. In the future it will be tested in reality, optimized, and improved. Indeed, there is room for improvements and additional developments as, for instance, adding I.A.-based systems to rank pairs, using machine learning and deep learning to set more precise entry levels for the positions. As a final consideration, according to what we mentioned in chapter 2, we would state that a more thorough test of cointegration, like Johansen's one that relies on trading sets of stocks instead of pairs, would be worth to be explored and could represent a more powerful and sophisticated evolution of the strategy to guarantee more robust returns with lesser risk exposition.

Bibliography

Y. Amihud. Illiquidity and stock returns: cross-section and time-series effects. *Journal of financial markets*, 5(1):31–56, 2002.

Z. Bodie et al. *Investments*. Tata McGraw-Hill Education, 2009.

R. F. Engle and C. W. Granger. Co-integration and error correction: representation, estimation, and testing. *Econometrica: journal of the Econometric Society*, pages 251–276, 1987.

E. F. Fama and K. R. French. *Common risk factors in the returns on stocks and bonds*. University of Chicago Press, 2021.

E. Fischer and S. Lind-Braucher. Optimal portfolios with traditional and alternative investments: An empirical investigation. *The Journal of Alternative Investments*, 13, 12 2009. doi: 10.3905/jai.2010.13.2.058.

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

R. Hyndman and G. Athanasopoulos. *Forecasting: Principles and Practice*. OTexts, Australia, 2nd edition, 2018.

F.-S. Lhabitant. *Handbook of hedge funds*, volume 332. John Wiley & Sons, 2007.

L. Pástor and R. F. Stambaugh. Liquidity risk and expected stock returns. *Journal of Political economy*, 111(3):642–685, 2003.

G. Vidyamurthy. *Pairs Trading: quantitative methods and analysis*, volume 217. John Wiley & Sons, 2004.