University of Warsaw

Faculty of Economic Sciences

Zhao Wang

Student's book no.: 328857

CO-INTEGRATION IN PAIR TRADING STRATEGIES.
Application to high-frequency data on American equities

Second cycle degree thesis
Field of study: Finance and Accounting
Specialty: Quantitative Finance

The thesis written under the supervision of
Dr. Piotr Wojcik
Chair of Economic Development Theory
Faculty of Economic Sciences UW

Warsaw, January 2015

Oświadczenie kierującego pracą

Oświadczam, że niniejsza praca została przygotowana pod moim kierunkiem i stwierdzam, że spełniła ona warunki do przedstawienia jej w postępowaniu o nadanie tytułu zawodowego.

Data 27.01.2015

Podpis kierującego pracą

Statement of the Supervisor on Submission of the Thesis

I hereby certify that the thesis submitted has been prepared under my supervision and I declare that it satisfies the requirements of submission in the proceedings for the award of a degree.

Date 27.01.2015

Signature of the Supervisor

Oświadczenie autora (autorów) pracy

Świadom odpowiedzialności prawnej oświadczam, że niniejsza praca dyplomowa została napisana przeze mnie samodzielnie i nie zawiera treści uzyskanych w sposób niezgodny z obowiązującymi przepisami.

Oświadczam również, że przedstawiona praca nie była wcześniej przedmiotem procedur związanych z uzyskaniem tytułu zawodowego w wyższej uczelni.

Oświadczam ponadto, że niniejsza wersja pracy jest identyczna z załączoną wersją elektroniczną.

Data

2015-01-27

Podpis autora (autorów) pracy

Statement of the Author(s) on Submission of the Thesis

Aware of legal liability I certify that the thesis submitted has been prepared by myself and does not include information gathered contrary to the law.

I also declare that the thesis submitted has not been the subject of proceedings resulting in the award of a university degree.

Furthermore I certify that the submitted version of the thesis is identical with its attached electronic version.

Date

2015-01-27

Signature of the Author(s) of the thesis

王杨

Summary

The aim of this research is to develop a profitable high frequency pair trading strategy based on selected stocks of S&P 500. The analysis is performed with co-integration method on the 1-minute high-frequency data covering the period Jan 1, 2010 - Dec 31, 2012. Empirical analysis presented in research indicates that the pairs from the same industry performed better than the pairs from different industry and the weak-form Efficient Market Hypothesis (EMH) does not hold here. The best-performed out-of-sample net Sharpe Ratio (SR) is 0.41 in financial sector.

Celem pracy jest stworzenie zyskownej strategii inwestycyjnej typu pair-trading operującej na danych wysokiej częstotliwości dla wybranych spółek z indeksu S&P 500. W strategii wykorzystywana jest analiza kointegracji na danych jednominutowych z okresu od 2010-01-01 do 2012-12-31. Analiza empiryczna prezentowana w pracy wskazuje, że wybrana strategia inwestycyjna zastosowana na parach spółek z tego samego sektora daje lepsze wyniki niż strategia zastosowana na parach pochodzących z różnych sektorów oraz że słaba forma hipotezy rynków efektywnych (EMH) w tym przypadku nie jest spełniona tylko w przypadku spółek z sektora finansowego, w których uzyskano w okresie out-of-sample net Sharpe ratio równe 0.41.

Key words

< Pair-trading，High-frequency trading, investment strategy, Co-integration approach >

Area of study (codes according to Erasmus Subject Area Codes List)

<14.3 Economics>

Theme classification: Economics, Finance

The title of the thesis in Polish
Kointegracja w strategii pair-trading. Zastosowanie do danych wysokiej częstości dla amerykańskich akcji

Table of contents

Introduction

In 1980s a stock-trading group from Morgan Stanley explored trading opportunities by applying the quantitative methods. The members of that trading group included professionals in physics, math and computer science. This trading group led by Nunzio Tartaglia made a significant amount of profit for company in 1987 by applying a simple quantitative approach "pairs trade". Since then this trading strategy has gained more and more attention in industry till now. In general, the pair trading is to trade stocks with similar historical price trend. Specifically go short with the overvalued stock and go long undervalued one at the same time when the spread deviates from the historical mean. The pairs trading strategy can be applied in various assets, not only in stocks, but also ETFs, futures, etc. The paper by Dunis.C.L, Giorgioni.G (2010) "Statistical Arbitrage and High-Frequency Data with an Application to Eurostoxx 50 Equities" indicates that it can be applied in high-frequency data. The pair trading requires that prices of a pair of stocks have a stable long-term relationship, which allows for making arbitrage from the short-term mispricing. The concept of co-integration can make a good indication of long-term relationship between two or more data series. The aims of the research include: build a co-integration based pair trading strategy and test a hypothesis that if the weak form EMH does not hold here. The 1-minute high frequency data for stocks included in S&P 500 index is used. The sector approach applied in Golonkiewicz-Rybska.Kaja's research "Pairs trading strategies on US equities" contributes insights to this paper: the GICS (Global Industry Classification Standard) sector classification is applied in stock sector classification process, the stocks from same sector are assumed to be most likely to move in the same direction. Transaction cost is also considered. It varies for different kinds of securities, one can typically estimate it by taking half the average bid-ask spread of a security and then add the commission. The transaction cost applied in this paper is assumed on the level of 15 basis points (0.15%); a round trip transaction (entry and exit from position) will therefore cost 30 basis points in this case.

The strategy can be simplified as follows: for all pairs considered (within and between sectors) the co-integration test is applied pair, (the Engle-Granger two-step method will be used as a testing approach) on a daily basis. If co-integration is found for the data of a particular day the pair is traded on the next day with a simple volatility breakout strategy. The profit & loss (P&L) and Sharpe Ratio will be used as performance measures. This research will allow verifying the performance of the strategy and also provide a basic idea for further research with application of more sophisticated tools.

The remaining part of the thesis is structured in the following way. Chapter 1 is giving a brief introduction of history of pair-trading strategy and the general review of related significant studies in this field. It sets out the general structure of a pair trading strategy in real financial market. In chapter 2 we explains main methods applied in this study, co-integration approach is the main tool applied in pair trading and it enables us to detect the long-run relationship between two variables, which is fundamental in pair trading. The SR is main comparison ratio to measure our strategy performance. Chapter 3 explains the details of the data applied in this study and strategy framework of a pair trading strategy, besides that there are two hypotheses is to be tested here. The empirical results will be presented in chapter 4, the aggregated results for all pairs from same sector to different sector and 6 different sectors respectively allow us to find the answers to hypotheses.

Chapter 1. Pair trading - historical perspective and literature review

This chapter presents the basic knowledge about pair trading strategy and its related historical background. The insight of strategies is based on reviewing major literature researches in academic field and pragmatic reports from related financial market practitioners.

1.1 Pair trading as an example of a trading strategy

There are four different types of traders out there in financial market: the investor is a type of a trader who just wants to hold securities for long periods; the market-maker makes profit from the bid-ask spread; arbitrageur gains from mispriced financial assets, sells the overpriced and buys underpriced then profits from the difference; the predictor uses data analysis techniques to make prediction about future price changes. The trading strategy is the key of any trading company, most of trading strategies are kept secretly, but some of them are just obvious.

In general when there are obvious signs that the security price is mean reverting or trending, it indicates that there exist some profitable trading strategies. The rule "Buy low and then sell high if a current price is low relative to the reference price" could generate profits from a mean reverting strategy. "If the price is trending and currently low, short sell then buy at an even lower price later" would also make profits. Academic research indicates that stock prices are on average very close to random walking. But it doesn't prove that under some special conditions they cannot exhibit some degree of mean reversion or trending behavior. The pair trading strategy is developed based on the behavior of a stock price series mean reversion. During the pair trade, the pair portfolio is bought when the spread of the stock prices formed by theses pairs is low, and sold/short when is high. In other words, this is a classic mean-reverting strategy. The stock price series is "stationary" if it never drifts farther away from its initial value, and that would be a great candidate for a mean-reversion strategy. Unfortunately, most stock price series are not stationary. However, it sometimes appears for a pair of stocks that after taking long-short position the market value of the pair is stationary. In this case, the two series are said to be co-integrated. More details about the concept of co-integration will appear in the methodological part.

Typically, two stocks that form a co-integrating pair are from the same industrial sector. Traders have long been familiar with this so-called "pair-trading strategy". And the pair trading is a type of market-neutral strategy. The trading strategy is considered market-neutral if it seeks to offset market risk, it is done by hedging. The market-neutral strategy is not a single strategy. It includes: convertible arbitrage, equity hedge, fixed-income arbitrage, relative value arbitrage and so on. Pair trading strategy is applied by taking long short position and hedge against market risk. When two stock prices move into the same direction in long-term period, it creates a profit opportunity. For instance, A and B are two stocks highly related and after testing the long-term movement by some statistical tools, when A goes up and B goes down, the profit can be made by shorting A and going long in B if they revert to the mean.

In fact, not all the pair traders are familiar with the concepts of stationarity and co-integration. Correlation is frequently used, but it's not same thing as co-integration. Correlation between two stock price series actually indicates correlating relationship of their returns over some time period, when there is positive correlation it means that the two series will move in the same direction in most time. However the positive correlation does not say anything about the long-term behavior of the two stocks. More important, it does not ensure that two stock price series will not grow farther and farther apart in long-term period even if they do move in the same direction on most days. However, if two stocks were co-integrated, the price series will be unlikely to diverge; yet their daily (or weekly, monthly) returns may be uncorrelated.

A profitable strategy is developed only if the stock prices are either mean-reverting or trending and the transaction costs also play a huge part of role in trading. There is a three-step process in pair trading strategy design in real world: identification of stock pairs, co-integration testing, and trading rule formulation. In the following chapters this will be described in more details.

1.2 History of pairs trading & literature review

The initial idea of pair trading starts from the legendary Wall Street trader Jesse Livermore in 1980s. He invented a trading strategy called “sister stocks”. He choosed the stocks from the same industry, which prices had some degree of correlation. Then he went short with the over-performing one and long with the weak performing one at the same time and closed the position when the price came back to the normal mean. It’s a classic market-neutral strategy, it hedges most market risk and the correlation between strategy returns and general market return is rather low.

In 1985, Dr Tartaglia started a team that focused on researching profitable quantitative trading strategy, it made 50 million dollars profit after two years trading. But his so called “pair-trading” performed weak in the following years, and then the team was dismissed. Although it only lasted for a short period, it inspired more individual traders as well as some big institutions in industry to apply this simple trading strategy making profits. The profit of this strategy is decreasing due to the secret leaks - more and more people applied this simple strategy to trade in market affect the profit in general. Dr Tartaglia considered the pair trading strategy as a psychological approach. He claimed that people usually buy stocks after they go up not down. One of the techniques used in trading is to identify if two stocks potentially move in same direction. If the pair relationship was detected, whenever the series is changing abnormally, the pair would finally correct itself. Two stocks are highly correlated, the one stock goes up and another one goes down, the trader will go long the underperforming one and go short the over-performing one at the same time. The profit is made if the traders believe the pair will revert to the mean finally. The correlation degree between two stocks is measured by the correlation coefficient. But the correlation coefficient does not promise that the pair will move in the same direction in long term, so the co-integration approach is adopted by the more and more pair traders. If two stocks are said to be co-integrated, so the pair of stocks will move in the same direction in long term.

The common idea behind pairs trading is short the overvalued securities and long the undervalued ones. But it's not easy to find the true value of the security in reality in order to judge whether is overvalued or undervalued. The concept of relative pricing may solve this problem, two stocks have been involved in similar business, and then the price must be similar. If the prices are different, one is the overpriced, the other one is underpriced, or both of them are mispriced. The mispricing between two stocks is captured by the notion of a spread. The greater the spread, the greater the potential profit. The returns from a pair trading strategy are uncorrelated to market returns; it's a typical feature of a market neutral strategy. By taking matching long and short position one essentially creates a hedge, where one trade acts as a hedge against the other. By doing this trade, the risk is controlled. Even in the financial crisis, the prices fall for both stocks. The losses from long position will be mitigated by the gain of the short position.

In reviewing literature period, besides look through most major pair trading research papers in field, the reports from financial practitioners are also contributed great amount of knowledge of pair trade in real finance word. The general method applied in this study follows the common rule of exploring a pair of stocks that moves together then taking matching long-short position. Jacobs.Bruce.I, Levy.N.Kenneth (1993) categorize long/short equity strategies as market-neutral, equitized, and hedge strategies. The market-neutral strategy holds longs and shorts in equal dollar balance at all times. This approach eliminates net equity market exposure, so the returns provided should not be affected by the market's direction. The equitized strategy, besides holding longs and shorts then adds a permanent stock index futures overlay in an amount equal to invested capital. The hedge strategy also holds long short position in equal dollar but has a variable equity market exposure based on market outlook. The exposure is hedged by using stock index futures.

Do, Robert and Hamza (2006) build the different methods to long/short position strategy, the difference between the strategies originates from the definition of "mispricing", but both require to open long/short positions at the same time.

Do, Robert and Hamza (2006) conclude there are three main methods to implement pairs trading: the distance method, the co-integration method and the stochastic spread method. The distance method captures the "co-movement" in a pair of normalized stock series, it executes trades when the sum of square differences reaches a certain threshold. Traders normally select the pair of stocks that has minimum distance. Upper/lower trading band is assumed on the level of two historical standard deviations.

Elliott.J.R (2005) captures mean-reverting behavior of pair stocks. They models mean reversion of price spread. The price spread is denoted as X and follows a Vasicek process: dx_t=k $(\theta - x_t)\ d_t + \sigma dB_t$, where dB_t is a standard Brownian motion and the X is going to revert to the mean θ at speed κ. By making the spread equal to $y_t = x_t + Hw_t$, the pair-traders declare that the observed spread is driven mainly by a mean-reverting process, plus some measurement error where $w_t \sim \mathcal{N}\ (0,1)$. This stochastic spread method offers advantages from empirical perspective: it is able to model the mean-reverting behavior that underlies pairs trading, it is convenient for forecasting and the model is easy-to-use (Kalman Filter can easily estimate the parameters in this setting). Unfortunately there is a fundamental issue in this approach: in long-term, the pair of stocks is restricted to yield the same return such that any departure from it will be expected to be corrected in the future. It's not easy to find a pair of stocks that generate same return in reality. However, in this study we focus on co-integration method.

The co-integration method presented in Vidyamurthy.G (2004) is an attempt to parameterize pairs trading, by exploring the possibility of co-integration (Engle and Granger, 1987). The traditional statistical tool for portfolio optimization is correlation analysis of asset returns, but the co-integration approach has been more and more popular in recent years. Alexander and Dimitriu (2002) point that correlation is only efficient for stationary variables, it is done by prior de-trending of prices and other variables, which are found to be integrated of order one or higher. But this process will cause loosing important information. However the co-integration analysis is able to detect any stochastic trend of price and then use these trends for a dynamic analysis of correlation in returns (Alexander.C, Giblin.I, Weddington.W III 2002). Compared to correlation, co-integration is allowed to use the whole information of financial variables, and it indicates long-run relationship of co-integrated pairs. It requires to be integrated of order one when use in stock price. If a pair of stocks is co-integrated, then its spread is mean reverting and the mean reversion is the most important feature of pair trading strategy.

Chapter 2. Methodology

2.1 Concept of co-integration

The co-integration is an econometric property that indicates the long-term relationship between two (or more) data series. If two or more series are non-stationary, but a linear combination of them is stationary, then the series are said to be co-integrated. Stationary means that the mean and variance do not change over time and do not follow any trends. The mathematical definition of "stationary process": let $\{X_t\}$ be a stochastic process and let $F_X(X_{t1+\tau}, \dots, X_{tk+\tau})$ represent the cumulative distribution function of the joint distribution of $\{X_t\}$ at times $t_1+\tau$,..., $t_k+\tau$. Then, $\{X_t\}$ is said to be stationary if, for all K, for all τ, and for all t_1,..., t_k,

$$F_X(X_{t1+\tau}, \dots, X_{tk+\tau}) = F_X(X_{t1}, \dots, X_{tk}). \quad (1)$$

Since τ does not affect $F_X(.)$, F_X is not a function of time.

In the definition, the time series X_t and Y_t are co-integrated of order d, b, where d$\geq b > 0$, what is expressed as:

$$X_t,\ Y_t \sim \mathrm{CI}\ (\mathrm{d}, \mathrm{b}) \quad (2)$$

if both time series are integrated of order d, and there exists a linear combination of those variables, $a_1x_t+a_2y_t$, that is integrated of order d - b. The vector $[a_1,\ a_2]$ is called a co-integration vector. What's more, many economic and stock price series are integrated of order 1, or I (1). A stationary series is by definition an I (0) process.

The main methodology applied in this research in order to detect stocks that move together involves concept of co-integration, which was developed by Engle and Granger (1987), it also concerns two or more variables. The main methodology of testing it applied in this research is the Engle-Granger two-step method. By this approach it allows to detect if two or more financial variables are integrated of the same order then estimate the co-integrating vector by using OLS regression. In the end one has to test the obtained residuals, check whether they are stationary. Engle-Granger approach indicates that if time series X_t and Y_t are co-integrated, a linear combination of them must be stationary. This can be represented in a formula:

$$\mu_t = y_t - \beta x_t \tag{3}$$

Where x_t and y_t are non-stationary variables, μ_t is stationary. After obtaining μ_t, the famous test approaches such as ADF (Augmented Dickey Fuller), PP (Phillips-Perron) and KPSS (Kwiatowski-Phillips-Schmidt-Shin) could be applied to testify the stationarity. The β is estimated first by applying OLS (Ordinary Least Squares), after then testify the stationarity. The co-integration will be tested for every pair within a particular industry and across industries. If the pair exhibits co-integration on a particular day, then it will be traded on the following day under some other condition related to price spread. The details of the strategy will be described in the following chapter. It allows for selecting viable pairs and choosing best performing set of parameters in the in-sample periods. Then strategy performance will be back-tested on the out-of-sample period. The strategy performance will be analyzed in the empirical part.

2.2 Unit root test

Unit root test is the most common way to test if the stock series variable is stationary or not. There are three main methods which are commonly accepted by the practitioners in the industry: Augmented Dickey-Fuller test, Phillips-Perron test and

Kwiatkowski-Phillips-Schmidt-Shin test. In this kind of approach one has to firstly set up the existence of a unit root as the null hypothesis, and then check the critical value whether the result indicates the hypothesis is rejected or not. Unit root test is used to test whether there exists an absolute value of root which equals to 1 in a characteristic polynomial equation of auto-regression model. Consider the following AR model:

$$X_t = \sum_{i=1}^{p} \varphi_i X_{t-i} + \varepsilon_t \tag{4}$$

Where φ_i is model parameter, ε_t is random error. It can be also presented using a lag operator:

$$\varphi(L)X_t = \varepsilon \tag{5}$$

Where $\varphi\,(L) = 1 - \sum_{i=1}^{P} \phi\, L^P$ is polynomial of lag operator. According to the formula representations above, the unit root test can be presented as following a hypothesis testing question:

$$H_0: \varphi(:) = 0, \ni |:| = 1 \leftrightarrow H_1: \varphi(:) \neq 0, \forall |:| = 1 \tag{6}$$

AR model has a unit root if unit root test results cannot reject the null hypothesis; then it indicates that the time series is not stationary time series. The **Augmented Dickey-Fuller test** is an augmented version of the "Dickey-Fuller Test". The Dickey-Fuller test checks whether a unit root is present in an autoregressive model. The unit root test is able to test the long-term relationship between the time series variables. The test is named after statisticians David Dickey and Wayne Fuller, who developed the test in 1979.

The hypothesis test of ADF presents as follows:

$$H_0: \delta=0 \ \leftrightarrow \ H_1: \delta \neq 0 \tag{7}$$

Different variants of ADF test can be used in practical application: simply assuming that the series includes a unit root $\Delta X_t = \delta X_{t-1}$+$\nu_t$; testing a unit root with drift $\Delta X_t = \mu + \delta X_{t-1}$+$\nu_t$, where μ is a drift; Test unit root with a drift and a deterministic time trend $\Delta X_t = \mu + \beta t + \delta X_{t-1} + \nu_t$, where μ drift, βt is deterministic time trend among $\nu_t = -\sum_{i=1}^{p} \alpha_i \Delta X_{t-i} + \varepsilon_t$. Compared to Dickey-Fuller, ADF is an augmented version which removes all the structural effects / autocorrelation in the time series and then tests using the same procedure. The ADF test statistic, used in the test, is a negative number. The more negative it is, the stronger rejection of the hypothesis that there is a unit root at some level of confidence. If the value of tau-statistic is higher than the 5% critical value -2.86, the null hypothesis about non-stationarity cannot be rejected, which in co-integration test means no co-integration.

Kwiatkowski–Phillips–Schmidt–Shin (KPSS) test is another way to test a null hypothesis with a deterministic trend. For instance, let a deterministic time trend, a stationary residual and a random walk represent series $Y_1, Y_2, \dots, Y_N$ where t is the time index, μ_t is independent identically distributed.

$$Y_t = \beta t + (r_t + \alpha) + e_t \qquad (r_t = r_{t-1} + \mu_t \text{ is a random walk}) \tag{8}$$

The null hypothesis: Y_t is level stationary; the alternative hypothesis: Y_t is a unit root process. The Lagrange Multiplier test can test the hypothesis. If in empirical analysis the value of test statistic is higher than 5% critical value, then the null hypothesis about stationarity is rejected. The formula given by:

$$\text{KPSS} = T^{-2} \sum_{T-1}^{T} \hat{S}_t^2 / \hat{\lambda}^2 \tag{9}$$

$$\hat{S}_t^2 = \sum_{T-1}^{T} \hat{u}_j \tag{10}$$

Phillips - Perron Test is another unit root test applied also in testing for co-integration process. It is used to test the null hypothesis that data series is integrated of order 1. It builds on the DF test of the null hypothesis.

$$X_t = \rho X_{t-1} + \varepsilon_t \tag{11}$$

The null hypothesis of PP test given as follows:

$$H_0: \rho = 1 \;\leftrightarrow\; H_1: \rho \neq 1 \tag{12}$$

The PP is able to test a transformative version of AR model:

$$X_t = \alpha + \beta t + \rho X_{t-1} + \varepsilon_t \tag{13}$$

Where α means a drift of AR model; βt indicates the trend related to time in AR model.

2.3 Measuring performance of the strategy

The size of risk has significant impact on the portfolio performance under the theory of modern portfolio investment. The rate of return after adjusting can be considered as a good measurement between risk and return. It allows to eliminate the negative impact of risk factor to performance evaluation. The **Sharpe Ratio** is a classic measurement of return to risk ratio. It's widely adopted by the practitioners in financial industry due to its ability to consider risk factor and return at the same time. In general, the higher the expected return on assets, the higher the ability of risk tolerance; the lower expected return, the lower ability of risk tolerance. The rational investors usually prefer to maximize the profit under the risk-taking limitation; minimize the risk under the fixed profit plan. The Sharpe Ratio is developed by Nobel laureate William Sharpe. The theory is based famous CAPM (Capital Asset Pricing Model) to measure the performance of financial assets. The main concept of

Sharpe Ratio: investors require the return rate equal or greater than the risk-free rate when they plan to invest in something risky.

$$SR\text{=[E } (E_P) \text{ - } R_f] \text{ / } \sigma P \tag{14}$$

Where E (R_P) is return rate of portfolio; R_f is risk free rate; σp is standard deviation. The formula indicates that the how much profit can be generated from taking every 1 more unit of risk. The original concept of return rate comes from Capital Market Line. It has been considered as one of the most common rate for measurement. The Sharpe Ratio is an adequate measure for risky financial assets investment, SR represents the profits from undertaking every one more unit of risk. The positive number indicates the average return rate higher than the volatile risk, the negative figure indicates that the risk greater than the return. By SR it is possible to calculate the scale of risk and return in portfolio, the greater figure indicates better performing portfolio. For example: the return rate of Treasury Bill is 3% and the expected return is 15%, the standard deviation of portfolio is 6%, the SR equals 2 (12%/6%) which indicates the extra 2% will generate from undertaking 1% risk. In high frequency trading with no positions carried overnight, the position carrying costs are 0; the high-frequency SR:

$$\text{SR} = \frac{\bar{r}}{\sigma_r} \tag{15}$$

In general if the average and standard deviation of returns is based on a certain trading period T (a month, a day, or an hour) and the practitioners usually to annualize theses quantities, firstly need to find out how such trading periods there are in whole year (N_t).

$$\text{Annualized Sharpe Ratio} = \sqrt{N_t} \times \text{ Sharpe Ratio Based on T} \tag{16}$$

SR is easy-to-use in terms of calculation, but when it applies in reality there are some disadvantages listed by financial practitioners. SR gains the popularity but it cannot capture the tail risk of extreme adverse returns and there also is a group of researchers present studies against using SR because of non-normally distributed returns. SR method ignores deviations from normality may lead to under-estimate risk and over-estimate performance. Normally the negative returns are significant when it comes to estimate the performance of strategies, but SR may happen to occur positive returns in volatility measure. The number of SR ratio has no significance and it's used only in comparison. Ratio is linearly related, but the relationship between risk and return is not linear, there is room of biased error when take SR as fund measurement tool. Despite all those disadvantages, the SR is still widely adopted by the financial industry due to its simplicity and doesn't require overmuch assumption.

Jensen's Alpha is also a measure of performance that extracts from broad market influence, CAPM-style, it takes into consideration the variability of returns in co-movement with chosen market indices.

$$\propto_i = \text{E}\ [r_i]\ -\ r_f\ -\ \beta_i(r_M - r_f) \tag{17}$$

Where β_i is the regression coefficient of trading returns on portfolio, such as the market portfolio. It measures the trading return in excess of the return predicted by CAPM. It is suitable only if returns are normally distributed and the investor wishes to split his holdings between one trading strategy and the market portfolio. Another measure ratio developed by **Treynor** is a risk-adjusted measure of return based on systematic risk. It is very similar to SR, the difference is in the fact that the **Treynor Ratio** uses beta as the measurement of volatility.

$$Treynor_i = \frac{E\ [r_i] - r_f}{\beta_i} \tag{18}$$

Where β_i is the regression coefficient of trading returns on portfolio, such as the market portfolio. The ratio suits if returns are normally distributed and the investor wishes to split his holding between one trading strategy and the market portfolio. These methods we described is widely used in industry, but they share a common disadvantage that they are not able to capture the tail risk of extreme adverse returns. Recently some new performance measures have been subsequently developed to capture the tail risk inherent in the returns of most trading strategies. There is an improved version of Sharpe ratio that enables to capture the tail risk of trading strategies. It is done by changing the measure of risk from standard deviation to a drawdown-based methodology. Calmar ratio and Burke ratio also work well in capturing tail risk. Calmar ratio (Young, Terry 1991) uses the maximum drawdown as the measure of volatility:

$$\text{Calmar}_i = \frac{E[r_i] - r_f}{-MD_{i1}} \tag{19}$$

where MD_{i1} denotes the lowest maximum drawdown. The Burke ratio (Burke, 1994) uses the standard deviation of maximum drawdowns as a volatility metric:

$$Burke_i = \frac{E[r_i] - r_f}{\sqrt{\sum_{K=1}^{N} MD_{ij}^2}} \tag{20}$$

where MD_{i1} denotes the lowest maximum drawdown, MD_{i2} second lowest maximum drawdown and so on (here: a type of variance below the N_{th}largest drawdown in the denominator - accounts for very large losses). The concept of maximum drawdown indicates the maximum severity of losses observed based historical data and it is the difference between the global maximum with the minimum of the equity curve after the occurrence of previous maximum peak point. A drawdown then can be described as the lowest return in between two continuous maximum peak points.

Chapter 3. Data Description and analysis framework

The data applied in this research comes from www.quantquote.com and includes 1 minute data of 100 most liquid stocks from S&P 500 Index. The analyzed period is from 1 January 2010 to 31 December 2012. The standard & Poor 500 is a stock price index that records 500 largest companies based on market capitalization. The S&P 500 covers the most companies listed in US major stock exchange, like NYSE, NASDAQ. Compared to Dow Jones, S&P 500 contains more companies so that the risk is more diversified, it's better to reflect the market change. S&P 500 was invented by Standard & Poor Corporation in 1957 and initially started from 425 different stocks. It sets a base index from 1941 to 1942 and is calculated as a weighted average. Compared to Dow Jones, S&P 500 contains more industries with more precise data. It's considered as an ideal stock index. The GICS (Global Industry Classification Standard) will be applied in this research, which is a classification of industry developed by MSCI and S&P for financial use. It consists of 10 different sectors, 24 industry groups, 67 industries based on market value of most major public companies. The table 1 shown below indicates the top 100 companies of S&P 500 index categorized into GICS industrial sectors.

Table 1. GICS taxonomy for top 100 S&P 500 constituents

Sector	**Amount of firms**	**Industry Groups**
Energy	9	Energy
Materials	5	Materials
Industrials	11	Capital Goods
		Commercial & Professional Services
		Transportation
Consumer Discretionary	16	Automobiles & Components
		Consumer Durables & Apparel
		Hotels Restaurants & Leisure
		Media
		Retailing
Consumer Staples	12	Food & Staples Retailing
		Food, Beverage & Tobacco
		Household & Personal Products
Health Care	14	Health Care Equipment & Services
		Pharmaceuticals & Biotechnology
Financials	15	Banks
		Diversified Financials
		Insurance
		Real Estate
Information Technology	13	Software & Services
		Technology Hardware & Equipment
		Semiconductors & Semiconductor Equipment
Telecommunication Services	2	Telecommunication Services
Utilities	3	Utilities

Source: own elaboration based on GICS.

3.1 Data Description

The data from analyzed period is from 1 January 2010 to 31 December 2012 is divided into 12 quarters. I decided to use every second quarter as the in-sample period and leave the remaining ones for additional model evaluation. The detailed structure of in-sample, out-of-sample period is presented below:

Table 2. In/out sample allocation

TIME	QUARTER	PERIOD ASSIGNMENT
2010	1	In-sample
	2	Out-of-sample
	3	In-sample
	4	Out-of-sample
2011	5	In-sample
	6	Out-of-sample
	7	In-sample
	8	Out-of-sample
2012	9	In-sample
	10	Out-of-sample
	11	In-sample
	12	Out-of-sample

Source: Own elaboration.

The raw data of 100 most liquid stocks from S&P 500 will be categorized into 10 different sectors in this research. Due to the massive calculation process, for the purpose of this research we will finally use the data for top 5 stocks (based in market capitalization) from 6 major industries. Therefore finally data for 30 companies were analyzed included in the following sectors: Information Technology, Energy, Industrial, Consumer Staples, Financials and Health Care. The list of considered stocks ordered with respect to capitalization in each of 6 sectors analyzed is presented in table 3. The finally selected top 5 stocks from each sector are shaded in gray.

Table 3. Top 5 stocks among 6 major sectors

Sector.	NYSE.	Cap	Sector.	NYSE.	Cap
INFORMATION TECHNOLOGY	AAPL	674	***INDUSTRIALS***	GE	269
	MSFT	406		UNP	107
	GOOG	365		MMM	102
	ORCL	182		UTX	98
	INTC	165		UPS	96
	IBM	162		BA	92
	CSCO	136		HON	75.8
	QCOM	119		CAT	62
	MA	96		DHR	58
	EBAY	68		EMR	44
	EMC	62		DE	32
	TXN	55	***HEALTH CARE***	JNJ	303
	V	8		PFE	192
CONSUMER STAPLES	WMT	270		MRK	170
	PG	237		GILD	151
	KO	191		AMGN	121
	PEP	147		BMY	97
	PM	135		UNH	93
	CVS	103		CELG	83
	MO	97		LLY	75
	WAG	64		BIIB	71
	CL	62		MDT	69
	COST	60.1		ABT	66
	KMB	42		ESRX	58
	KFT	34		BAX	39
FINANCIALS	BRK.B	358	***ENERGY***	XOM	403
	WFC	276		CVX	217
	JPM	225		SLB	122
	BAC	180		COP	88
	C	162		OXY	67
	AXP	93		EOG	53
	GS	82		APC	45
	USB	78		HAL	41
	AIG	75		APA	27
	MET	62			
	SPG	56			
	PNC	46			
	BK	45			
	COF	45			
	AMT	40			

Source: Own elaboration.

There will be 30 stocks in total, 5 from each of 6 major industries, it indicates there will be 435 possible combinations / pairs. The following step is to check co-integration in each pair by applying the basic two-step Engle-Granger procedure. The two-step Engle-Granger procedure consists of running a regression between variables and then testing for the stationarity of residuals. Co-integration indicates the long-term relationship and once found it will be logical to apply it in pair-trade. Co-integration test is applied in 6 sectors for all 435 pairs on the rolling basis - for every single day (so every test is based on 390 observations of 1-minute prices of the whole day of quotations between 9:30 and 16:00). Table 4 set out the 30 stocks applied in this study, it shows the full names of companies and the industrial sector which they belong to.

Table 4. The list of 30 stocks applied in pair-trading strategy

No.	NYSE.	Full name	Sector
1	KO	The Coca-Cola Co	CONSUMER
2	PEP	Pepsico, Inc.	CONSUMER
3	PG	The Procter & Gamble	CONSUMER
4	PM	Philip Morris International	CONSUMER
5	WMT	Wal-Mart Stores Inc.	CONSUMER
6	COP	ConocoPhillips	ENERGY
7	CVX	Chevron Corporation	ENERGY
8	OXY	Occidental Petroleum	ENERGY
9	SLB	Schlumberger Limited	ENERGY
10	XOM	Exxon Mobil Corporation	ENERGY
11	BAC	Bank of America	FINANCIAL
12	BRK_B	Berkshire Hathaway	FINANCIAL
13	C	Citigroup Inc.	FINANCIAL
14	JPM	JPMorgan Chase & Co	FINANCIAL
15	WFC	Wells Fargo & Company	FINANCIAL
16	AMGN	Amgen Inc.	HEATLTHCARE
17	GILD	Gilead Sciences Inc.	HEATLTHCARE
18	JNJ	Johnson & Johnson	HEATLTHCARE
19	MRK	Merck & Co. Inc.	HEATLTHCARE
20	PFE	Pfizer Inc.	HEATLTHCARE
21	GE	General Electric	INDUSTRIAL
22	MMM	3M Company	INDUSTRIAL
23	UNP	Union Pacific Corporation	INDUSTRIAL
24	UPS	United Parcel Service	INDUSTRIAL
25	UTX	United Technologies Corporation	INDUSTRIAL
26	AAPL	Apple Inc	IT
27	GOOG	Google Inc	IT
28	INTC	Intel Corporation	IT
29	MSFT	Microsoft Corporation	IT
30	ORCL	Oracle Corporation	IT

Source: Own elaboration.

3.2 Strategy Framework

The technique entry is key to trading strategies, the good entry is one that triggers a trade at the moment of low risk and high return. Breakout entry techniques are simple and intuitively appealing: the security is bought when prices break above an upper band, sold when prices break below a lower band. In this study we will apply the volatility breakout model because it is more contemporary and sophisticated. The trading trigger long/short positions are based on volatility bands, as volatility increase, the bands expand and move farther away from current price; as it decreases, the band come closer to the market. It was efficient before the advanced computational power became cheap and popular, nowadays simple breakout methods cannot guarantee to generate good profits. Anyway, it's still a good method for long-term trading, it could be practical if the parameter setting is appropriate. The spread in this research is calculated as P_1- P_2*m, where m=m_1/m_2 is based on average ratio between the prices on the previous day. Spread is a signal to our model, which shows whether to take position or not (volatility band around spread). For comparison reason we set out different multipliers m = 1.5, 2, 2.5, 3, 3.5. The different parameter will give us different results, it will help to choose best parameter setup in in-sample period in order to get the optimal outcome.

Strategy of this research can be simplified as follows: build 435 possible combinations based on 30 stocks selected from 6 major industries. (AAPL&GOOG, AAPL&INTC… XOM&WFC). In the second step for each pair detect on a daily basis the strength of relationship by correlation, linear regression and co-integration respectively. Build a spread between the two stocks with the spread multiplier based on the average ratio of prices on the preceding day. Trade the pair on the following day only if the relationship is strong enough (correlation above the level of 0.7 in absolute sense, significant linear regression and significant co-integration) by applying a simple volatility breakout model using rolling standard deviations with different memories: 30, 60 and 90, 120 minutes. The entry techniques applied in this research is a volatility breakout models. The pair is sold when the spread breaks above an upper band or threshold, it's bought short when the spread breaks below a lower band or threshold. The volatility breakout model is applied with different volatility multipliers: m = 1, 1.5, 2, 2.5, 3, 3.5. The transaction cost also has significant impact on the performance of portfolios, in this research it is assumed on the level of 15 bps (0.15% one way), SR and PnL measure the performance of a portfolio. We apply this strategy to 435 pairs respectively, and then aggregate the final results. With the help of Dr Wojcik, the loop of aggregation in R save us a great amount of time and energy in order to get final results, the aggregated results of all pairs in/out sample with 1- 3.5 volatility band and 30-120 minutes rolling window size are measured by gross/net SR and gross/net PnL based co-integration, regression and correlation. The automated chart generated from aggregation of raw results helps us to easily find out the best-performed parameter set-up in order to apply it in out-of-sample period.

3.3 Research hypotheses

The main aim of this research is to check whether the strategy we designed is profitable. There exists a number of successful trading strategies based on the concept of pair trading. However the performance declines because of more and more investors applied this kind of strategy since it leaks in 90s. Before we move to empirical analysis, it is logical to set up the formula of hypotheses for this study. Two hypotheses will be verified here:

1. EMH: Efficient market hypothesis does not hold in a weak form - it is possible to create a consistently profitable trading strategy using only historical quotations.

The main goal of this research is build up a pair trading strategy then test whether it is profitable, which may contradict the classic theory of EMH, efficient markets hypothesis. Eugene Fama developed EMH in 1970. The strong, semi strong and weak form efficiency of EMH is distinguished. The strong form insists that the price contains all the information of company operation, it contains public and non-public. There is no way to make extra profit in the market. Semi-strong form believes that the price reveals all the public information about company operation; all the technical analysis won't be profitable except the insider trading. The weak form can be tested in research, all the technical analysis method won't be profitable, fundamental analysis may lead investor profit from market. In recent years EMH has more controversies due to the assumption that both fundamental and technical analysis have no sense; only higher risk taking can generate higher revenue. The hypothesis in this research indicates that the weak form of EMH does not hold.

2. The portfolios composed of pairs from same sector perform better than the portfolio of pairs from different sectors.

The pairs are formed based on the relationship between two stocks assuming they will move in the same direction, so it is logical to believe that the stocks in the same industries will have similar trend and they have greater possibility that they will move in the same direction. The aggregate results generated from the analysis will enable us to compare the profit/loss of pairs from the same sector and different sectors.

Chapter 4. Empirical Results

4.1 In-sample evaluation:

In this research, we examine the results based on co-integration, correlation and regression methods and then compare them with respect to different aggregation methods: For all possible pairs, all possible pairs from the same sectors; all possible pairs in the different sectors; all pairs from the same sector for each sector separately.

The raw data of top 100 constituents of S&P 500 index are divided into 10 sectors, and then we choose top 5 stocks from each of 6 major sectors to test our hypotheses. The strategy is built mainly based on co-integration approach. The co-integrated relationship is found by testing the data on a daily basis. If the long-term relationship is found, we then apply the volatility breakout model. The results are generated in 6 different sizes of volatility brand in order to find the best parameters. Sharpe Ratio is the main indicator to determine the performance of the strategy; only average values of gross SR and net SR calculated on all in-sample or all out-of-sample quarters are presented.

(1) Summary for all pairs in the same sector:

In order to compare the performance between same/different sectors and test our hypotheses, we summarize the SR of all pairs from the same sector, and there are 10 pairs in each sector, so 60 pairs totally. We aggregate the results based on co-integration, correlation and regression methods with 30…120 different memories of standard deviation and 1…3.5 different volatility multipliers.

Table 5. Aggregated in-sample results for pairs in same sector

AvQ.net SR													
M	**Coint**				**Cor**				**Reg**				**Average**
	30	**60**	**90**	**120**	**30**	**60**	**90**	**120**	**30**	**60**	**90**	**120**	
1	-6.4132	-5.9675	-5.6418	-5.4559	-19.3525	-17.1877	-15.4692	-14.3023	-34.2630	-28.1839	-24.5841	-21.7223	*-16.55*
1.5	-5.6222	-5.1546	-4.7291	-5.1073	-16.1941	-17.1877	-15.4692	-11.6557	-26.4060	-28.1839	-24.5841	-17.4412	*-14.81*
2	-5.2940	-4.6882	-4.5465	-4.5323	-14.2867	-12.1776	-10.7810	-10.2915	-22.0983	-18.2734	-16.2620	-14.9619	*-11.52*
2.5	-4.9695	-4.2688	-4.0865	-4.2568	-12.9344	-10.9860	-9.2596	-8.9076	-19.6985	-16.5518	-14.1962	-13.0645	*-10.27*
3	-4.7449	-4.1203	-4.1358	-4.1533	-12.1466	-10.4067	-9.2784	-8.7035	-18.2043	-15.4424	-13.6426	-12.5248	*-9.79*
3.5	-4.5573	-3.9191	-3.8196	-3.7822	-11.3179	-9.6921	-8.9377	-8.2286	-16.9053	-14.2502	-12.9820	-11.7217	*-9.18*
Average	*-5.27*	*-4.69*	*-4.49*	*-4.55*	*-14.37*	*-12.94*	*-11.53*	*-10.35*	*-22.93*	*-20.15*	*-17.71*	*-15.24*	/
AvQ.gross SR													
M	**Coint**				**Cor**				**Reg**				**Average**
	30	**60**	**90**	**120**	**30**	**60**	**90**	**120**	**30**	**60**	**90**	**120**	
1	-0.3211	-0.2557	-0.5060	-0.5741	3.0940	2.9624	2.8588	3.0668	3.7935	3.6169	3.4343	3.4763	*2.05*
1.5	-0.5139	-0.4684	-0.5550	-0.9333	2.6973	2.2364	2.2627	2.5219	3.1287	2.8528	2.7987	2.6849	*1.56*
2	-0.5701	-0.6116	-0.7256	-0.7055	2.4107	1.7623	2.2426	2.1037	2.6212	2.3830	2.4587	2.3212	*1.31*
2.5	-0.5803	-0.5587	-0.5266	-0.6508	2.2854	1.7540	2.3461	2.3221	2.3116	2.0777	2.4656	2.6924	*1.33*
3	-0.5424	-0.5476	-0.6743	-0.4396	2.1308	1.8532	2.1717	2.4750	2.1013	2.0379	2.1158	2.4776	*1.26*
3.5	-0.5172	-0.4574	-0.3601	-0.3276	1.8908	1.7976	2.1878	2.7716	1.9108	2.1614	2.2105	2.6662	*1.33*
Average	*-0.51*	*-0.48*	*-0.56*	*-0.61*	*2.42*	*2.06*	*2.34*	*2.54*	*2.64*	*2.52*	*2.58*	*2.72*	/

Note: Values in column headers (30, 60, 90, 120) indicate different memories of standard deviation and values in row headers (1, 1.5, 2, 2.5, 3, 3.5) are different volatility multipliers used in the breakout model.

Source: Own elaboration

The table above presents an aggregated version of summary results, it indicates the performance of all pairs in the same sectors. SR is the main indicator for the strategy performance. The results present net and gross values separately in order to consider the impact on transaction cost (15 bps). The strategy is built mainly based on co-integration method, but we also add correlation and regression in order to compare the different results based on different method. We based our strategy on correlation, regression and co-integration: regression is always extremely significant, so in case of this filtering method each pair is trading on every single working day; for correlation we assume the threshold on the level of 0.7. If the relationship between two series is confirmed, then apply the volatility breakout model. We apply rolling analysis of standard deviation with different size of rolling window (recent 30, 60, 90, 120 minutes of the spread observation). We also setup different value of the volatility multiplier parameter for comparison: different parameter for upper bound and lower bound (1, 1.5, 2, 2.5, 3, 3.5). The results indicate different performance and it helps to determine which combination we may choose as a final strategy. In gross terms (without transaction costs), the table indicates the best-performed strategy is based on regression method with volatility band 1, rolling window size for standard deviation equal to 30 minutes. The average gross SR for this variant is equal to 3.7935 and it is the best figure of all considered here. In the co-integration based strategy the best performing variant is the one with volatility band equal to 3 and rolling window of standard deviation equal to 60 minutes. It's gross SR equals -0.2557. After adding the transaction costs, the results show differently - no strategy is giving profits which are able to cover transaction costs: the best-performed strategy is the one based on co-integration with volatility band 3.5 and rolling window of standard deviation equal to 120 minutes - it's net SR equals -3.7822. With the consideration of cost, we can determine the strategy based on co-integration method is well performing in general, although the figures are still not positive. Despite the transaction cost, the increase in volatility multiplier leads to lower benefit based on average values presented in above table, and the increase in memories of standard deviation leads to higher benefit in general, although there are parts of the table, where numbers behave adversely. When we take into consideration transaction costs, the increase in volatility multiplier leads to higher outcome and the increase in memories of standard deviation leads to higher benefit. We observe that without taking into consideration the transaction cost, the results based on co-integration are negative while for correlation and regression positive. So co-integration is not a good filtering method here, the regression and correlation works better here. This may be caused by the fact that co-integration is rather rare phenomenon in the tested data and more frequent trading (based on regression and correlation filter) allows for making higher profits. However, the transaction costs has significant impact on our performance, after we take into consideration of transaction cost, the co-integration approach remains the least negative SR.

(2) Summary for all pairs in the different sectors

We summarized all pairs from different sectors, there are 375 pairs in total (total number of pair combinations 435 minus 60 pairs from same sector). Table 6 presents aggregated in-sample results for pairs in different sectors. The second way of aggregation gives similar result compared to the previous situation under consideration of transaction costs. The strategy based on co-integration with 3.5 volatility band and rolling window of standard deviation equal to 120 minutes performed best in terms of SR ratio (net) = -6.0301. In gross terms the best one is the strategy based on regression method with 1 volatility band and rolling standard deviation based on recent 30 minutes - it's SR gross equals 5.2679. Here we can test the hypothesis that the pairs from the same sector are performing better than the pairs from different sectors - we can compare them in terms of SR indications. Table 7 presents the difference in net SR (just for the filtering based on co-integration) between results of all pairs from the same sector and all pairs from different sectors. All the figures are positive, which is proves the second research hypothesis that in the co-integration framework the pairs from the same sector perform better than the pairs from different sectors.

In gross terms, all the numbers shows positive but the regression based method performed best. After we take into consideration the transaction cost, co-integration based approach remains the least negative. Despite the transaction cost, the increase in volatility multiplier leads to lower benefit based on average value presented in table, and the increase in memories of standard deviation leads to lower benefit in general, although there parts of the table where numbers behave adversely. When we take into consideration the cost, the increase in volatility multiplier leads to higher outcome and the increase in memories of standard deviation leads to higher benefit although they'll are all negative.

Table 6. Aggregated in-sample results for pairs in different sectors

AvQ.net SR													
M	**Coint**				**Cor**				**Reg**				**Average**
	30	**60**	**90**	**120**	**30**	**60**	**90**	**120**	**30**	**60**	**90**	**120**	
1	-16.7684	-14.4936	-12.6004	-11.2724	-16.5519	-14.7596	-13.3926	-11.9643	-38.0544	-31.0030	-26.4937	-23.0909	-19.20
1.5	-13.3730	-11.1325	-9.6649	-8.8555	-14.1628	-12.1972	-10.7739	-9.8382	-29.2454	-23.9869	-20.5672	-18.3939	-15.18
2	-12.5087	-9.2905	-8.1664	-7.4289	-12.5087	-10.6629	-9.1333	-8.4253	-24.6504	-20.3012	-17.4330	-16.1736	-13.06
2.5	-10.0939	-7.9226	-6.9886	-6.5203	-11.4787	-9.6888	-8.2819	-8.0464	-22.1624	-18.2753	-16.0069	-15.2418	-11.73
3	-8.9959	-7.1819	-6.7883	-6.5711	-10.7886	-9.1177	-7.9550	-7.8526	-20.5461	-17.1557	-15.0712	-14.6838	-11.06
3.5	-8.3073	-6.5328	-6.2713	-6.0301	-10.1525	-8.5064	-7.7246	-7.6903	-19.2630	-16.1052	-14.5109	-14.1875	-10.44
Average	-11.67	-9.43	-8.41	-7.78	-12.61	-10.82	-9.54	-8.97	-25.65	-21.14	-18.35	-16.96	/

AvQ.gross SR													
M	**Coint**				**Cor**				**Reg**				**Average**
	30	**60**	**90**	**120**	**30**	**60**	**90**	**120**	**30**	**60**	**90**	**120**	
1	2.5582	2.4306	2.4484	2.3884	4.4836	4.3165	4.2980	4.1521	5.2679	4.9676	4.7847	4.6012	3.89
1.5	2.5876	2.3600	2.3893	2.1708	4.1447	3.9730	3.9729	4.1512	4.6162	4.3516	4.4249	4.2278	3.61
2	2.3268	2.3106	1.9621	2.1095	3.8646	3.7713	3.7782	3.9236	4.0880	3.9821	3.9105	3.8581	3.32
2.5	2.1140	2.3099	2.1423	2.2217	3.6200	3.5774	3.6788	3.7455	3.6539	3.6483	3.6164	3.6420	3.16
3	2.1441	2.2693	2.0029	2.0606	3.4455	3.3803	3.3449	3.4828	3.3465	3.3685	3.3816	3.3743	2.97
3.5	2.1740	2.3225	2.2218	2.3885	3.3157	3.4068	3.2313	3.4101	3.1682	3.3475	3.2206	3.2743	2.96
Average	2.32	2.33	2.19	2.22	3.81	3.74	3.72	3.81	4.02	3.94	3.89	3.83	/

Note: Values in column headers (30, 60, 90, 120) indicate different memories of standard deviation and values in row headers (1, 1.5, 2, 2.5, 3, 3.5) are different volatility multipliers used in the breakout model.

Source: Own elaboration

Table 7. Difference in net SR for in-sample period (same sector - different sectors)

AvQ.net SR (Same - Different sector)					
M	Co-integration				Average
	30	60	90	120	
1	10.36	8.53	6.96	5.82	*7.91*
1.5	7.75	5.98	4.94	3.75	*5.60*
2	7.21	4.60	3.62	2.90	*4.58*
2.5	5.12	3.65	2.90	2.26	*3.49*
3	4.25	3.06	2.65	2.42	*3.10*
3.5	3.75	2.61	2.45	2.25	*2.77*
Average	*6.41*	*4.74*	*3.92*	*3.23*	/

Note: Values in column headers (30, 60, 90, 120) indicate different memories of standard deviation and values in row headers (1, 1.5, 2, 2.5, 3, 3.5) are different volatility multipliers used in the breakout model.

Source: Own elaboration

(3) Summary for pairs from the same sector for each sector separately:

Consumer staples sector:

In order to compare the performance sector by sector, we apply the analysis through all the sectors separately. The consumer staples sector contains 5 companies: KO (Coca Cola), PEP (Pepsi), PG (Procter & Gamble Co), PM (Philip Morris International Inc.), WMT (Wal-Mart Stores Inc.). We apply our strategy designed in this research into consumer staples sector; the result is in line with general results for all pairs presented before. The strategy based on co-integration with 120 rolling window size and 3.5 volatility band generates the best-performed net SR. Despite the transaction cost, the increase in volatility multiplier leads to lower benefit based on average value presented in table, and the increase in memories of standard deviation leads to higher benefit in general, when we take into consideration the cost, the increase in volatility multiplier leads to higher outcome and the increase in memories of standard deviation leads to higher benefit. We observe that without taking into consideration the transaction cost, the results based on co-integration are negative while for correlation and regression are positive. So co-integration is not a good method for filtering days of trading for this sector, the regression and correlation works better. However, the transaction costs still has impact on performance, after we take into consideration the transaction cost, the co-integration approach remains the least negative SR.

Table 8. Aggregated in-sample results in Consumer Staples sector

AvQ.net SR													
M	**Coint**				**Cor**				**Reg**				**Average**
	30	**60**	**90**	**120**	**30**	**60**	**90**	**120**	**30**	**60**	**90**	**120**	
1	-5.2688	-5.1019	-4.7160	-4.6357	-12.2314	-11.5387	-10.8548	-9.9219	-29.7480	-26.1272	-22.8638	-20.4503	*-13.62*
1.5	-5.1055	-4.6503	-4.3746	-3.8779	-11.1879	-9.7957	-8.8307	-7.6829	-24.9668	-20.5419	-17.5155	-15.4784	*-11.17*
2	-4.8435	-4.4984	-4.0737	-3.7886	-10.0290	-8.9716	-7.9387	27.0000	-21.3051	-17.6190	-15.3180	-13.4338	*-7.07*
2.5	-4.4140	-4.2492	-3.8178	-3.7005	-9.0779	-8.0502	-6.7908	-6.0758	-19.0258	-15.6209	-13.0972	-11.9335	*-8.82*
3	-4.1530	-4.0318	-3.7167	-3.4038	-4.0318	-7.2441	-5.9549	-5.3482	-17.4559	-13.9333	-11.5264	-10.9565	*-7.65*
3.5	-4.2024	-3.6483	-3.7718	-3.2601	-3.6483	-6.4193	-5.7014	-4.8522	-16.4085	-12.5777	-10.6253	-10.3304	*-7.12*
Average	*-4.66*	*-4.36*	*-4.08*	*-3.78*	*-8.37*	*-8.67*	*-7.68*	*-1.15*	*-21.49*	*-17.74*	*-15.16*	*-13.76*	/

AvQ.gross SR													
M	**Coint**				**Cor**				**Reg**				**Average**
	30	60	90	120	30	60	90	120	30	60	90	120	
1	-0.8705	-0.5557	-0.6779	-0.5704	2.5651	2.4854	2.3902	2.3804	2.9453	2.8876	2.8907	2.8229	*1.56*
1.5	-0.7177	-0.6650	-0.4962	-0.0362	2.3923	1.6488	1.9966	2.4707	2.7784	1.7888	2.0313	2.6222	*1.32*
2	-0.6141	-0.6985	-0.4815	-0.3564	1.9622	1.7098	2.0351	2.0190	2.3243	1.8041	2.4703	2.3788	*1.21*
2.5	-0.6141	-0.6985	-0.4815	-0.3564	1.9622	1.7098	2.0351	2.0190	2.3243	1.8041	2.4703	2.3788	*1.21*
3	-0.6196	-0.6571	-0.3057	-0.1192	1.7071	1.6202	2.3300	2.3226	1.9270	2.0252	2.6622	2.3467	*1.27*
3.5	-0.7213	-0.5443	-0.7780	-0.0822	1.5765	1.8955	2.0704	2.1808	1.8221	2.2997	2.6230	2.2899	*1.22*
Average	*-0.69*	*-0.64*	*-0.54*	*-0.25*	*2.03*	*1.84*	*2.14*	*2.23*	*2.35*	*2.10*	*2.52*	*2.47*	/

Note: Values in column headers (30, 60, 90, 120) indicate different memories of standard deviation and values in row headers (1, 1.5, 2, 2.5, 3, 3.5) are different volatility multipliers used in the breakout model.

Source: Own elaboration

Energy Sector:

The energy sector contains 5 companies as well: COP (ConocoPhillips), CVX (Chevron Corporation), OXY (Occidental Petroleum Corporation), SLB (Schlumberger Limited), XOM (Exxon Mobil Corporation). The results are similar as for the previous sector, the less frequent trading based on co-integrated relationship generates higher net SR (although negative). The general pattern of results is that highest net SR appears in the strategy based on co-integration method with higher volatility band and higher rolling window size. Obviously, the trading strategy we designed seems not to make profits after including the cost, but this research can help us to test the hypothesis. In gross term, the results based on co-integration are negative while for correlation and regression are positive. So co-integration is not a good method of filtering days of trading also for this sector. The regression and correlation works better. However, the transaction costs still has impact on performance, after we taking into consideration of transaction cost, the co-integration approach remains the least negative SR.

Table 9. Aggregated in-sample results in energy sector

AvQ.net SR													
M	**Coint**				**Cor**				**Reg**				**Average**
	30	**60**	**90**	**120**	**30**	**60**	**90**	**120**	**30**	**60**	**90**	**120**	
1	-4.4240	-4.1454	-3.9681	-3.8130	-15.9599	-14.1496	-12.6464	-11.4913	-21.6421	-18.2178	-15.9524	-14.4403	*-11.74*
1.5	-4.0403	-3.3570	-3.1842	-2.9004	-13.4309	-11.2184	-9.8970	-8.7081	-17.1792	-13.9603	-12.5048	-11.0519	*-9.29*
2	-3.7485	-3.3478	-2.5402	-2.3760	-11.5482	-9.4237	-8.2848	-7.2467	-14.4650	-11.7594	-10.3034	-9.1369	*-7.85*
2.5	-3.5674	-3.1013	-2.4806	-3.1271	-10.3083	-8.3262	-7.1376	-6.6111	-12.8763	-10.3916	-8.9053	-8.0371	*-7.07*
3	-3.4956	-2.9908	-2.4681	-2.9028	-2.9908	-7.3850	-6.4629	-5.8951	-11.7753	-9.3263	-8.1233	-7.4496	*-5.94*
3.5	-3.1610	-2.8048	-2.4051	-2.7086	-2.8048	-6.7754	-6.1754	-5.4468	-10.7699	-8.5997	-7.4580	-6.8578	*-5.50*
Average	*-3.74*	*-3.29*	*-2.84*	*-2.97*	*-9.51*	*-9.55*	*-8.43*	*-7.57*	*-14.78*	*-12.04*	*-10.54*	*-9.50*	/

AvQ.gross SR													
M	**Coint**				**Cor**				**Reg**				**Average**
	30	60	90	120	30	60	90	120	30	60	90	120	
1	0.0855	-0.0015	-0.1331	-0.2369	2.8851	2.7584	2.8374	2.8687	2.9932	2.8946	2.8674	2.7948	*1.88*
1.5	0.0203	-0.1000	0.1854	0.3216	2.5823	2.4360	2.5242	2.7629	2.7155	2.6150	2.4533	2.7637	*1.77*
2	0.1226	-0.3972	0.3707	0.1500	2.3591	2.0364	2.4619	2.7746	2.4856	1.9817	2.4781	2.8688	*1.64*
2.5	0.1639	-0.2702	0.2040	-0.2891	2.1398	2.1200	2.3749	2.4738	2.1566	2.0123	2.3792	2.7293	*1.52*
3	-0.3084	-0.3835	0.1847	-0.1786	2.0101	2.2920	2.3740	2.6613	2.1294	2.1989	2.4001	2.8948	*1.52*
3.5	-0.1303	-0.3819	0.4823	-0.1706	-0.3819	2.3767	2.1028	2.7752	2.1053	2.2661	2.4233	3.0441	*1.38*
Average	*-0.01*	*-0.26*	*0.22*	*-0.07*	*1.93*	*2.34*	*2.45*	*2.72*	*2.43*	*2.33*	*2.50*	*2.85*	/

Note: Values in column headers (30, 60, 90, 120) indicate different memories of standard deviation and values in row headers (1, 1.5, 2, 2.5, 3, 3.5) are different volatility multipliers used in the breakout model.

Source: Own elaboration

Financial Sector:

In financial sector we find: BAC (Bank of America Corp), BRK_B (Berkshire Hathaway Inc.), C (Citigroup Inc.), JPM (JPMorgan Chase & Co), WFC (Wells Fargo & Co). In gross terms, all the numbers shows positive, but the regression based method performed best. After we taking into consideration the transaction cost, co-integration based approach remains the least negative results. Despite the transaction cost, the increase in volatility multiplier leads to lower benefit based on average values presented in table, and the increase in memories of standard deviation leads to higher benefit in general. When we take into consideration the cost, the increase in volatility multiplier leads to higher outcome and the increase in memories of standard deviation leads to higher benefit although they'll are negative. The approach based on co-integration with 60 minutes standard deviation and 2.5 volatility multiplier has the least negative SR over all in net term. The results follow the general pattern we described before.

Table 10. Aggregated in-sample results in financial sector

AvQ.net SR													
M	**Coint**				**Cor**				**Reg**				**Average**
	30	**60**	**90**	**120**	**30**	**60**	**90**	**120**	**30**	**60**	**90**	**120**	
1	-2.8880	-2.5379	-2.2028	-2.0020	-10.2699	-8.9462	-7.7624	-6.8571	-13.5199	-10.8487	-9.3103	-8.0243	*-7.10*
1.5	-1.9966	-0.9514	-0.6293	-0.8713	-8.0166	-6.9024	-5.5021	-4.9436	-10.0401	-8.1235	-6.3734	-5.5720	*-4.99*
2	-1.8162	-0.1548	-0.4598	-0.8012	-7.0133	-5.4643	-4.1482	-3.6538	-8.4400	-6.3272	-4.9086	-4.0775	*-3.94*
2.5	-1.5298	-0.1200	-0.5005	-0.4696	-5.9865	-4.4699	-3.4141	-3.3356	-7.1106	-5.0962	-3.7826	-3.2254	*-3.25*
3	-1.0489	-0.1963	-0.6600	-0.4085	-0.1963	-3.9071	-3.3616	-3.2083	-6.2072	-4.6259	-3.7187	-3.2303	*-2.56*
3.5	-0.8938	-0.2497	-0.1899	-0.4067	-0.2497	-3.4722	-3.3413	-3.0815	-5.5994	-3.9981	-3.5010	-3.1994	*-2.35*
Average	*-1.70*	*-0.70*	*-0.77*	*-0.83*	*-5.29*	*-5.53*	*-4.59*	*-4.18*	*-8.49*	*-6.50*	*-5.27*	*-4.55*	/
AvQ.gross SR													
M	**Coint**				**Cor**				**Reg**				**Average**
	30	**60**	**90**	**120**	**30**	**60**	**90**	**120**	**30**	**60**	**90**	**120**	
1	2.0545	1.9083	2.3011	1.7371	2.9636	2.6510	2.8296	2.7979	3.9862	3.7783	3.7570	3.6212	*2.87*
1.5	2.1442	2.2614	2.1905	1.6440	2.4499	2.0970	2.5489	2.5332	3.4269	3.0593	3.1266	3.1257	*2.55*
2	1.7215	2.3233	1.9560	1.5989	2.0924	2.2132	2.5392	2.8330	2.9277	2.9786	3.0987	3.2862	*2.46*
2.5	1.4764	2.0850	1.7551	1.3933	1.9811	2.4597	2.7364	2.3007	2.7197	2.9950	3.2905	3.0414	*2.35*
3	1.6724	2.0108	1.3304	1.8783	2.0055	2.5678	2.3629	2.0968	2.7136	3.0142	2.9056	2.6790	*2.27*
3.5	1.6876	1.7997	1.7038	1.7784	1.7997	2.5469	2.0683	2.0599	2.8305	3.0660	2.8040	2.5001	*2.22*
Average	*1.79*	*2.06*	*1.87*	*1.67*	*2.22*	*2.42*	*2.51*	*2.44*	*3.10*	*3.15*	*3.16*	*3.04*	/

Note: Values in column headers (30, 60, 90, 120) indicate different memories of standard deviation and values in row headers (1, 1.5, 2, 2.5, 3, 3.5) are different volatility multipliers used in the breakout model.

Source: Own elaboration

Health care Sector:

In health sector, there are: AMGN (Amgen, Inc.), GILD (Gilead Sciences, Inc.), JNJ (Johnson & Johnson), MRK (Merck & Co., Inc.), PFE (Pfizer Inc.). In gross term, all the numbers show positive, but the regression based method performed best in general. After we take into consideration the transaction cost, co-integration based approach remains the least negative results. Before the transaction cost, the increase in volatility multiplier leads to lower benefit based on average value presents in table, and the increase in memories of standard deviation leads to higher benefit in correlation and regression, but lower outcome in co-integration. When we take into consideration the cost, the increase in volatility multiplier leads higher outcome and the increase in memories of standard deviation leads to higher benefit although they are all negative. The approach based on co-integration with 120 minutes standard deviation and 3-volatility multiplier has the least negative SR over all in net terms. The results again follow the general pattern we describe before.

Table 11. Aggregated in-sample results in health care sector

AvQ.net SR													
M	**Coint**				**Cor**				**Reg**				**Average**
	30	**60**	**90**	**120**	**30**	**60**	**90**	**120**	**30**	**60**	**90**	**120**	
1	-4.7077	-4.2209	-3.7743	-3.3525	-12.4142	-11.2144	-10.3210	-9.2805	-25.1672	-21.4019	-19.0274	-16.7921	-11.81
1.5	-4.0571	-3.3468	-2.6506	-2.3101	-10.6320	-9.3799	-8.6210	-7.7406	-20.2144	-16.6833	-14.9199	-13.0735	-9.47
2	-3.4788	-2.4430	-2.1604	-1.6617	-9.6099	-8.0853	-7.4536	-6.6231	-17.4039	-14.0636	-12.8731	-11.2784	-8.09
2.5	-2.8546	-2.4957	-1.9348	-1.6972	-8.7811	-7.3433	-6.5770	-6.0222	-15.5758	-12.6578	-11.6585	-10.3149	-7.33
3	-2.3918	-2.0174	-1.8534	-1.5060	-2.0174	-7.0032	-6.3870	-5.3637	-14.4826	-11.9550	-10.7873	-9.4463	-6.27
3.5	-2.2900	-1.9257	-1.6615	-1.8023	-1.9257	-6.7128	-6.0159	-5.1799	-13.4535	-11.1709	-10.0195	-8.8961	-5.92
Average	-3.30	-2.74	-2.34	-2.05	-7.56	-8.29	-7.56	-6.70	-17.72	-14.66	-13.21	-11.63	/

AvQ.gross SR													
M	**Coint**				**Cor**				**Reg**				**Average**
	30	**60**	**90**	**120**	**30**	**60**	**90**	**120**	**30**	**60**	**90**	**120**	
1	1.2030	1.2182	1.1193	1.0994	1.3316	1.3133	1.2950	1.4797	1.3832	1.4956	1.5008	1.4603	1.32
1.5	1.2489	1.4739	1.3907	1.2159	1.3095	1.2183	1.0640	1.1268	1.3319	1.2555	1.2692	1.4249	1.28
2	1.4233	1.4241	1.2880	1.6196	0.9081	0.9095	1.0030	1.3757	0.9053	0.9497	0.9978	1.6505	1.20
2.5	1.5571	1.2196	1.0025	1.1437	0.7608	0.8365	1.1619	1.1513	0.6047	0.9133	1.0976	1.4566	1.08
3	1.5611	1.2076	0.9105	0.9175	0.5994	0.6995	0.8964	1.3643	0.4835	0.8383	0.9941	1.6222	1.01
3.5	1.3292	0.9276	0.8641	0.5203	0.9276	0.7466	0.9882	1.4525	0.5007	0.9103	1.1739	1.8294	1.01
Average	1.39	1.25	1.10	1.09	0.97	0.95	1.07	1.33	0.87	1.06	1.17	1.57	/

Note: Values in column headers (30, 60, 90, 120) indicate different memories of standard deviation and values in row headers (1, 1.5, 2, 2.5, 3, 3.5) are different volatility multipliers used in the breakout model.

Source: Own elaboration

Industrial Sector:

The Industrial sector includes: GE (General Electric Company), MMM (3M Co), UNP (Union Pacific Corporation), UPS (United Parcel Service, Inc.), UTX (United Technologies Corporation). In gross terms, all the numbers show to be positive but the regression based method performed best in general. After we take into consideration the transaction cost, co-integration based approach remains the least negative results. Before the transaction cost is included, the increase in volatility multiplier leads to lower benefit based on average value presented in the table, and the increase in memories of standard deviation leads to lower benefit in correlation and regression, but higher outcome in co-integration. When we take into consideration the cost, the increase in volatility multiplier leads to higher outcome and the increase in memories of standard deviation leads to higher benefit although they are all negative. The approach based on co-integration with 120 minutes standard deviation and 3.5-volatility multiplier has the least negative SR over all in net term.

Table 12. Aggregated in-sample results in industrial sector

AvQ.net SR													
M	**Coint**				**Cor**				**Reg**				**Average**
	30	**60**	**90**	**120**	**30**	**60**	**90**	**120**	**30**	**60**	**90**	**120**	
1	-4.5455	-4.3298	-4.1848	-3.8305	-12.1593	-10.7102	-9.4847	-8.4484	-22.1724	-18.8290	-16.3617	-14.2217	*-10.77*
1.5	-4.1073	-3.5720	-3.2079	-2.9518	-10.2367	-8.5159	-7.5004	-6.6053	-17.7293	-14.5256	-12.4942	-11.1893	*-8.55*
2	-3.6020	-2.8044	-2.7773	-2.5804	-8.6573	-7.1146	-6.0217	-5.5673	-14.5610	-11.7929	-9.9964	-9.3642	*-7.07*
2.5	-3.4571	-2.4505	-2.7281	-2.4127	-7.7570	-6.2446	-5.7460	-5.5511	-12.9249	-10.4087	-9.1675	-8.7101	*-6.46*
3	-3.1293	-2.3018	-2.3899	-1.7611	-2.3018	-5.5115	-5.7395	-5.4015	-11.6965	-9.3072	-8.5497	-8.3119	*-5.53*
3.5	-2.8341	-1.8172	-1.9189	-1.7191	-1.8172	-5.0901	-5.4341	-4.7931	-11.0210	-8.7665	-8.2223	-7.6564	*-5.09*
Average	*-3.61*	*-2.88*	*-2.87*	*-2.54*	*-7.15*	*-7.20*	*-6.65*	*-6.06*	*-15.02*	*-12.27*	*-10.80*	*-9.91*	*/*

AvQ.gross SR													
M	**Coint**				**Cor**				**Reg**				**Average**
	30	**60**	**90**	**120**	**30**	**60**	**90**	**120**	**30**	**60**	**90**	**120**	
1	0.5617	0.5531	0.1765	0.3568	3.5855	3.6507	3.5322	3.3533	4.3073	4.3865	4.1499	4.1162	*2.73*
1.5	0.1725	0.0956	0.2963	-0.0091	3.2506	3.0856	3.0119	3.0910	4.0430	3.7094	3.6606	3.4922	*2.32*
2	0.2551	0.1431	0.2932	0.0649	3.2981	2.9346	3.4019	2.9719	3.9923	3.6727	4.0103	3.5555	*2.38*
2.5	0.1709	0.4702	0.5223	0.4893	2.9758	3.0377	2.8797	2.5538	3.4251	3.5663	3.4802	3.2289	*2.23*
3	0.4825	0.7255	0.9961	1.1205	3.1319	3.1935	2.7374	2.2957	3.3083	3.5828	3.4409	3.1987	*2.35*
3.5	0.5299	1.0162	1.4427	1.4030	1.0162	3.2236	2.7269	2.7152	3.2186	3.3598	3.3274	3.4574	*2.29*
Average	*0.36*	*0.50*	*0.62*	*0.57*	*2.88*	*3.19*	*3.05*	*2.83*	*3.72*	*3.71*	*3.68*	*3.51*	*/*

Note: Values in column headers (30, 60, 90, 120) indicate different memories of standard deviation and values in row headers (1, 1.5, 2, 2.5, 3, 3.5) are different volatility multipliers used in the breakout model.

Source: Own elaboration

IT Sector:

The IT sector included: AAPL (Apple Inc.), GOOG (Google), INTC (Intel Corporation), MSFT (Microsoft Corporation), ORCL (Oracle Corporation). The strategy based on co-integration with 120 rolling window size and 3.5 volatility band generates the best-performing Net SR. In gross term, the increase in volatility multiplier leads to lower benefit based on average value presented in table. When we take into consideration the cost, the increase in volatility multiplier leads to higher outcome and the increase in memories of standard deviation leads to higher benefit. We observe that without taking into consideration of transaction cost, the results based on co-integration are negative while for correlation and regression are positive. So co-integration seems again not to be a good method here in gross term. However, the transaction costs still has impact on performance, after we take into consideration the transaction cost, the co-integration approach remains the least negative SR.

Table 13. Aggregated in-sample results in IT sector

AvQ.net SR													
M	**Coint**				**Cor**				**Reg**				**Average**
	30	**60**	**90**	**120**	**30**	**60**	**90**	**120**	**30**	**60**	**90**	**120**	
1	-3.8914	-3.6381	-3.3398	-3.3622	-11.9083	-10.2552	-9.1293	-8.4798	-19.6272	-16.0019	-13.9355	-12.4563	*-9.67*
1.5	-3.5382	-3.1751	-2.9400	-3.4539	-9.7121	-8.3936	-7.6325	-7.1044	-14.8953	-12.5023	-11.1154	-10.4805	*-7.91*
2	-3.2052	-2.9454	-3.0617	-3.0604	-8.4040	-7.2269	-6.5138	-6.4404	-12.4017	-10.4554	-9.5491	-9.2718	*-6.88*
2.5	-3.0513	-2.6685	-2.7892	-2.8664	-7.6697	-6.6598	-5.6366	-5.5212	-11.2007	-9.6203	-8.4869	-7.9962	*-6.18*
3	-2.9073	-2.6039	-2.7299	-2.7247	-2.6039	-6.3762	-5.7853	-5.4060	-10.3638	-9.0411	-8.2329	-7.6543	*-5.54*
3.5	-2.8078	-2.5358	-2.5020	-2.3740	-2.5358	-6.0574	-5.5281	-5.0839	-9.6782	-8.4187	-7.8721	-7.1775	*-5.21*
Average	*-3.23*	*-2.93*	*-2.89*	*-2.97*	*-7.14*	*-7.49*	*-6.70*	*-6.34*	*-13.03*	*-11.01*	*-9.87*	*-9.17*	*/*

AvQ.gross SR													
M	**Coint**				**Cor**				**Reg**				**Average**
	30	**60**	**90**	**120**	**30**	**60**	**90**	**120**	**30**	**60**	**90**	**120**	
1	-0.7326	-0.6530	-0.6676	-0.7904	1.4312	1.3364	1.2514	1.4820	1.9714	1.8350	1.6210	1.6937	*0.81*
1.5	-0.8714	-0.6379	-0.8272	-1.3058	1.1784	0.7968	0.7612	0.9629	1.4553	1.2563	1.1720	0.9634	*0.41*
2	-0.8609	-0.8091	-1.0666	-1.0260	0.9771	0.4808	0.7208	0.4657	1.1043	0.9991	0.8399	0.5516	*0.20*
2.5	-0.8014	-0.7519	-0.8778	-1.0129	1.0114	0.3922	0.8490	0.8786	0.9766	0.6631	0.8478	1.0430	*0.27*
3	-0.6958	-0.7563	-0.8912	-0.8655	0.7925	0.3600	0.6358	0.9049	0.7855	0.5156	0.5488	0.8240	*0.18*
3.5	-0.6408	-0.6790	-0.6355	-0.5509	-0.6790	0.3196	0.6983	1.1049	0.6340	0.6208	0.5926	0.9812	*0.15*
Average	*-0.77*	*-0.71*	*-0.83*	*-0.93*	*0.79*	*0.61*	*0.82*	*0.97*	*1.15*	*0.98*	*0.94*	*1.01*	*/*

Note: Values in column headers (30, 60, 90, 120) indicate different memories of standard deviation and values in row headers (1, 1.5, 2, 2.5, 3, 3.5) are different volatility multipliers used in the breakout model.

Source: Own elaboration

4.2 Summary for in-sample evaluation:

Table 14. In-sample best performed variants

(In-sample) Summary for well-performed SR:		
Category	**Net SR**	**Method**
All pairs from same sector	*-3.7822*	Coint-120, M=3.5
All pairs from different sector	*-6.0301*	Coint-120, M=3.5
Consumer staples sector	-3.2601	Coint-120, M=3.5
Energy sector	-2.3760	Coint-120, M=2
Financial sector	-0.1200	Coint-60, M=2.5
Health care sector	-1.5060	Coint-120, M=3
Industrial sector	-1.7191	Coint-120, M=3.5
IT sector	-2.3740	Coint-120, M=3.5

Source: Own elaboration

We summarize the best indicators from all the results, although net SR indicates that we lose after adding transaction costs (15 bps one way), but it allows us to test the hypothesis we argue in this research. From the figures, one can see again that the pairs from the same sector perform better than the pairs from different sectors; we can also conclude that the weak form of efficient-market hypothesis may hold here. The technical analysis can produce positive returns in gross terms in some cases, but they are not large enough to cover transaction costs. So the first research hypothesis of this thesis was verified negatively, while the second research hypothesis was verified positively.

4.3 Out-of-sample evaluation

After selecting relatively well-performing parameters for the strategy, we apply it into the out-of-sample data set. The net SR appears to be negative in all categories. The pair strategy we designed is not profitable; strategy applied to the all sectors is losing money. The strategy we designed in this research is not sophisticated enough to be applied in real financial markets, but the outcome shows some positive image. It indicates that after adding more techniques it may become a profitable strategy. In out-of-sample evaluation it proves that the pairs from same sectors performed on average better than the pairs from different sectors. The weak form of EMH holds also here in general - with consideration of transaction costs we cannot win with the market. The strategy we build shows positive SR in financial sector data in out-of-sample period and is just slightly negative in in-sample period, which indicates that with more technique improvement this simple strategy could be used for stocks from financial data. For further study one may add more advanced techniques in strategy-building process in order to get a more profitable version. We also got all the out-of-sample aggregated results in order to compare our original setup selection from in-sample, table 15 indicates that in out-of-sample period: the parameter setup of all pairs from the same sector, consumer staples sector, energy, financial and industrial sector has actually better results in terms of net SR, but the difference is not that significant and three groups/sectors is still the same. So our parameter method setup does still make sense in this research.

Table 15. Out-of-sample results and actual best-performed results/method

(Out-of-sample) Summary for well-performed SR:				
Category	**Net SR**	**Method**	***Best performed SR/Method***	
All pairs from same sector	*-3.28*	Coint-120, M=3.5	*-3.26*	Coint-120, M=3
All pairs from different sector	*-6.9*	Coint-120, M=3.5	** Same*	
Consumer staples sector	*-1.84*	Coint-120, M=3.5	*-1.83*	Coint-120, M=3
Energy sector	*-2.26*	Coint-120, M=2	*-1.55*	Coint-90, M=3.5
Financial sector	*-0.34*	Coint-60, M=2.5	***0.41***	Coint-120, M=3.5
Health care sector	*-1.26*	Coint-120, M=3	** Same*	
Industrial sector	*-2.2*	Coint-120, M=3.5	*-1.81*	Coint-120, M=3
IT sector	*-2.28*	Coint-120, M=3.5	* Same	

Source: Own elaboration

Conclusions

The high frequency data trading strategies have been adopted by more and more institutional investors and mature individual investors in recent years, due to the rapid development of computer technology. It's considered as a simple, effective and proved to be profitable. The basic idea of pair-trading is to find whether two stocks move in the same direction, and by taking long-short position to make profits. The main method applied in this research is co-integration. It indicates the long-term behavior between two financial assets. We test the co-integration relationship on a daily basis, if co-integration relationship is found we make relative position on next day. There were two different research hypotheses verified in this thesis: 1. Test if the EMH still holds by using just historical data in this trading strategy; 2. Check if the pairs from same sector perform better than the pairs from different sectors. In order to verify these hypotheses we applied 1-minute high frequency data of 30 stocks inform the S&P 500 index - top 5 stocks from each of 6 sectors considered: consumers, financials, energy, health care, industrial, IT. The number of 30 stocks applied in this pair-trading strategy allowed to generate a total number 435 possible pair combinations of all through all the sectors. According to theory the constituents in the same sector are assumed to behave similarly in general. The strategy framework is based on finding the relationship of co-integration, regression or correlation between a pair of stocks, and then taking relative position with different size of volatility band. The transaction costs were assumed on the level of 15 basis points one way. The net SR in in-sample period indicates that all the combinations are negative, but we still select the least negative one in order to get the verify it in out-of-sample period. The financial sector appears the only positive SR in terms of actual best-performed net SR in the out-of-sample period. The remaining sectors are still negative. Although most net SRs are negative and it indicates our strategy is not sophisticated enough to profit from the market, we tested our hypotheses in this research. The pairs from the same sector performed much better than the pairs from different sectors. The weak-form EMH does not hold here, we cannot win with market by applying such a simple technical approach.

List of tables

BIBLIOGRAPHY

Alexander.C, Giblin.I, Weddington.W III, *Co-integration and asset allocation: A new active hedge fund strategy,* Elsevier Science Ltd, 2002.

Alexander.C and Dimitriu.A, *The Co-integration Alpha: Enhanced Index Tracking and Long-Short Equity Market Neutral Strategies,* University of Reading, 2002.

Brown, Keith, Frank K. Reilly, *Analysis of Investments and Management of Portfolios,* Cengage Learning, P-941.

Dickey.A, Fuller.A, *Distribution of the Estimators for Autoregressive Time Series with a Unit Root*, Journal of the American Statistical Association, 1979.

Do.B, Faff.R, Hamza.Kais.H, *A New Approach to Modeling and Estimation for Pairs Trading*, Monashn University, May 29 2006.

Dunis.C.L, Giorgioni.G, Laws.J, Rudy.J, *Statistical Arbitrage and High-Frequency Data with an Application to Eurostoxx 50 Equities*, Liverpool John Moores University, March 2010.

Engle.F, Granger.J.W, *Co-integration and Error Correction: Representation, Estimation, and Testing,* Econometrica, Mar 1987.

Elliott.J.R, Hoek, Malcolm, *Pair trading*, University of Calgary, The University of Adelaide, The Australian National University, April 2005.

Gatev.E, Goetzmann.N.William, Rouwenhorst.G.K, *Pairs Trading: Performance of a Relative-Value Arbitrage Rule*, Boston College and Yale University, 2006

Golonkiewicz-Rybska.Kaja, *Pairs trading strategies on US equities. Sectoral high-frequency data approach,* University of Warsaw, August 2013.

Jensen.C, *The Performance of Mutual Funds in the Period 1945-1964*, Journal of Finance, 1968.

Jacobs.Bruce.I, Levy.N.Kenneth, *Long/Short Equity Investing*, Jacobs Levy Equity Management in Roseland (NJ 07068), 1993.

Kwiatkowski.D, Phillips.B, Schmidt, Shin.Y, *Testing the null hypothesis of stationarity against the alternative of a unit root,* Journal of Econometrics, 1992.

Phillips.B, Perron.P, *Testing for a Unit Root in Time Series Regression*, Biometrika, 1988.

Sharpe.William F, The Sharpe Ratio, The Journal of Portfolio Management, 1994.

Schmidt.D.Arlen, *Pairs Trading: A Co-integration Approach*, University of Sydney, November 2008.

Vidyamurthy.G, *Pairs Trading Quantitative Methods and analysis*, John Wiley&Sons Inc, 2004.

Young, Terry W, *Calmar Ratio: A Smoother Tool*, Futures magazine, 1991.