TALLINN UNIVERSITY OF TECHNOLOGY

School of Business and Governance

Department of Economics and Finance

Suvi Eräkorpi

REGRESSION TO THE MEAN PHENOMENON IN THE STOCK MARKET

Master´s Thesis

Supervisor: Lecturer Peeter Luikmel

Tallinn 2017

I declare I have written the master's thesis independently.
All works and major viewpoints of the other authors, data from other sources of literature and elsewhere used for writing this paper have been referenced.

Suvi Eräkorpi ……

(signature, date)

Student's code: 156476

Student's e-mail address: suvi.erakorpi@gmail.com

Supervisor Lecturer Peeter Luikmel:

The thesis conforms to the requirements set for the master's theses

……

(signature, date)

Chairman of defence committee:

Permitted to defence

……

(Title, name, signature, date)

TABLE OF CONTENTS

ABSTRACT

The concept of regression to the mean suggests that whenever participants are selected for the research because their scores on a measure are either very high or very low, they will tend to be less extreme on a second testing; that is, their scores will have regressed toward the mean (Graziano and Raulin 2004). The research evaluated the existence of regression to the mean phenomenon in the stocks of the Standard & Poor´s 500 index, and the possibility to gain abnormal returns with the use of the mean regression in the stock market. The author conducted a quantitative research to evaluate the existence of the regression to the mean phenomenon using yearly and monthly total returns as key ratios. The time frame of the research covered years between 2000 and 2015. The research was divided into four parts: mean regression in the average yearly total returns, mean regression in the average monthly total returns, decile movement analysis and regression analysis.

Key words: efficient market hypothesis, regression to the mean phenomenon, the S&P 500 index, random walk, mean regression

INTRODUCTION

The first time when the author came across with the regression to the mean phenomenon was while reading the book “Thinking, Fast and Slow” by Daniel Kahneman. The book presents regression to the mean phenomenon trough an example in comparison of the height of children to the height of their parents. The height of children tends to be always more mediocre than the height of their parents. Kahneman states that the phenomenon is common and can be detected in many different occasions. (2011, 179) The author wants to test whether the regression to the mean can be detected in the stock markets as well.

It is generally believed that securities markets are efficient in reflecting information about individual stocks and about the stock market. Thereby, neither technical analysis, which is the study of past stock prices to predict future prices, nor fundamental analysis, which is the analysis of financial information such as company earnings and asset values, to help investors to select undervalued stocks, would enable an investor to achieve returns greater than those that could be obtained by holding a randomly selected portfolio of individual stocks with comparable risk. (Malkiel 2003, 3)

If markets were efficient and no greater returns could be achieved, why is there so many professional investors and investment advisors on the markets, investors who spend their time to find undervalued stocks and try to do better than others. Some professors and researchers claim that markets are efficient and for example regression to the mean phenomenon is not valid in stock markets. On the other hand, others have studied regression to the mean phenomenon and say that it has been detected in the stock markets but considering the transaction costs and taxes, the gained return is minor. The final group of professors and researchers say that regression to the mean phenomenon can be detected in the stock market to make profit.

The objective of the research at hand, is to evaluate the existence of regression to the mean phenomenon in stocks of the Standard & Poor´s 500 index and evaluate whether is possible to gain

abnormal returns with the use of the mean regression in the stock market. To fulfil the given objectives, the following research questions will be answered:

- Can the regression to the mean phenomenon be detected in the monthly level?
- Can the regression to the mean phenomenon be detected in the yearly level?

The author believes to find mean regression from the stocks of the S&P 500 index. It is being tested whether the mean regression is significant enough to gain excess returns in the stock market with the phenomenon. To achieve the objectives of the thesis and to test hypothesizes, the author will implement a statistical analysis. The mean regression is a numerical phenomenon, and that is the reason why quantitative approach suits best for this thesis research. Quantitative methods emphasize objective measurements and the statistical, mathematical or numerical analysis of data collected through polls, questionnaires and surveys, or by manipulating pre-existing data. (University of Southern California, 2017)

The author will use the stocks in the Standard and Poor´s 500 index from the Standard and Poor´s index family for empirical analysis. The index contains 500 leading companies and their stocks, and captures approximately 80% coverage of available market capitalization. The index is widely regarded as the best single gauge of large-cap U.S equities. (S&P Dow Jones Indices, 2017b) It is designed to reflect the U.S. market and the U.S. economy. The time frame will cover years between 2000 and 2015. The author believes that 15 year-period is long enough to detect possible regression to the mean from monthly and yearly data. For statistical analysis, Microsoft Excel is used, and the analysis is split into four parts: mean regression in the average yearly total returns of the stocks, mean regression in the average monthly total returns of the stocks, decile dynamics analysis and regression analysis.

The first chapter of the thesis, the literature review, will give a general overview of the concept of the efficient market hypothesis, and how regression to the mean phenomenon fits into that contextual framework. Then, regression to the mean is presented through examples in everyday life and in the stock markets. The author will present views and ideas of scholars who have examined the phenomenon. It is then followed by an overview of the stock markets. A short description will contain history and development of the stock markets. The last part of the literature review is dedicated to the brief overview of the Standard & Poor´s 500 index. The chapter focuses

on the dynamics of the index and the eligibility criteria that are required from the companies that are added into the index. The second chapter of the thesis will introduce the methodological tools which the author will use in her research. Using the first chapter´s literature study as support, the author will conduct a statistical quantitative study. The author will provide information about data collection, data analysis methods and limitations of the research, just a few to mention. The third chapter consists of the results of the research and discussion. After discussion, some suggestions are provided by the author, and the research is summed up in the conclusions, where the main findings are summarized.

1. LITERATURE REVIEW

The author aims to study how much regression to the mean phenomenon has been analyzed in the past, which research methods have been used and what have been the results. The author also gives an overview to the topics related to the regression to the mean phenomenon. The literature review will consist of four main parts. The first part discusses the efficient market hypothesis and when markets are efficient, no arbitrage opportunity should be available to make abnormal returns. The second part will focus on the regression to the mean phenomenon by providing a short definition of the topic, examples of how the phenomenon is detected in both people lengths and in stock markets, just a few to mention. After that the literature review will continue with short descriptions of the stock markets and the Standard & Poor´s 500 index which is used as a base for the research. For the literature review, the author used both primary and secondary sources such as books, academic journals and various reports from institutes and organizations to gather information about the regression to the mean phenomenon and other topics related to that subject.

1.1 Efficient market hypothesis

The efficient market hypothesis was widely accepted already in the 1960s and 1970s by financial economists. It was generally believed that securities markets were efficient in reflecting information about individual stocks and about the stock market. The view was that when information arises, the news spread quickly and it is included into the prices of securities without any delay. Thereby, neither technical analysis, which is the study of past stock prices to predict

future prices, nor fundamental analysis, which is the analysis of financial information such as company earnings and asset values, to help investors to select undervalued stocks, would enable an investor to achieve returns greater than those that could be obtained by holding a randomly selected portfolio of individual stocks with comparable risk. (Malkiel 2003, 3)

Jensen (1978) defines market efficiency as follows: "A market is efficient with respect to information set, if it is impossible to make economic profits by trading on the basis of information set". A closely related definition of market efficiency is provided by Malkiel (1992): "A capital market is said to be efficient if it fully and correctly reflects all relevant information in determining security prices. Formally, the market is said to be efficient with respect to some information set, if security prices would be unaffected by revealing that information to all participants. Moreover, efficiency with respect to an information set, implies that it is impossible to make economic profits by trading on the basis of information set." Market efficiency can be divided into three forms based on how "all available information", the information set, is defined. If the information set includes past and current asset prices, the efficient market hypothesis is in its weak form. Expanding information set to include all publicly available information gives rise to the efficient market hypothesis in its semi-strong form. Finally, if all public and private information is included in the information set, market efficiency is in its strong form. (Granger and Timmermann, 2004)

The efficient market hypothesis is associated with the idea of a random walk. This term is used in the finance literature to characterize a price series where all subsequent price changes represent random departures from previous prices. The logic behind the random walk idea is that if the flow of information is unimpeded and information is immediately reflected in stock prices, then tomorrow´s price change will reflect only tomorrow´s news and will be independent of the price changes of today. Malkiel presents that if stock markets are efficient and totally unpredictable then a blindfolded chimpanzee throwing darts at the Wall Street Journal could select a portfolio that would do as well as the portfolio selected by experts. However, many financial economists and statisticians have begun to believe that stock prices are at least partially predictable, and patterns and certain valuation metrics can be created to predict future stock prices based on previous stock prices. (Malkiel 2003, 3-4)

Malkiel states that stock markets are far more efficient and far less predictable than some resent academic papers have showed earlier. He states that markets are more efficient when less

predictable patterns are available, and markets are efficient when they do not allow investors to earn above-average returns without above-average risks. In his article, The efficient market hypothesis and its critics, he presents several patterns of possible predictability that are used as an arguments against market efficiency, and refutes them. For example, predictable patterns based on valuation parameters such as predicting future returns from initial dividend yields have showed that investors have earned a higher rate of return from the stock market compared to buy and hold investment strategy but these findings are not necessarily inconsistent with efficiency. Even if some predictability exists, it may reflect time varying risk premiums and required rates of return for stock investors rather than an inefficiency. (Malkiel 2003, 16)

Another example is value stocks. There have been studies that suggest that value stocks have higher returns than so-called growth stocks. Samsa points out that in general, and consistent with the broader literature of behavioral finance, value-based strategies tend to outperform the market, but only a modest amount (2014, 107). The two most common methods of identifying value stocks have been price-earnings ratio and price-to-book-value ratio. Stocks with low price-earnings multiples appear to provide higher rates of return than stocks with high price-earnings ratios. The ratio of stock price to book value has also been found to be useful predictor of future security returns. Such results raise questions about the efficiency of the market but they may simply indicate failure of the capital asset pricing model to capture all the dimensions of risk. (Malkiel 2003, 19)

Even though, the main point of Malkilel´s article is that stock markets are efficient, he also points out two incidents that can be seen as cases of inefficiency by behaviorists. The first incident is the market crash of October 1987. Behaviorists say that the one-third drop in market prices, which occurred early in October 1978, can only be explained by psychological considerations since the basic elements of the valuation equation did not change rabidly over that period. Malkiel does not deny the role of psychological factors but in his opinion it would be a mistake to dismiss the significant changes in the external environment which can provide an entirely rational explanation for the significant decline in the values for common stock during the market crash. With the changes in external environment he is referring to the yields on long-term treasury bonds that increased from 9 percent to 10.5 percent in the two months prior to mid-October. In addition, there

were a few events that may have increased risk perceptions during the first two weeks of October. (Malkiel 2003, 25)

Another example of market inefficiency that Malkel is able to explain with rational reasons is the internet bubble of the late 1990s. It seemed that the valuations of the high-tech companies were proper. Many of the Wall Street´s most respected analysists were recommending Internet stocks to clients as being fairly valued. While it is now clear that such professionals were egregiously wrong, there was certainly no obvious arbitrage opportunity available. Thus, the stock market may well have temporarily failed in its role as an efficient allocator of equity capital, it does not mean that stock markets are not efficient as a whole. As long as stock markets exist, the collective judgement of investors will sometimes make mistakes. Undoubtedly, some market participants are more irrational than others. As a conclusion, pricing irregularities and predictable patterns in stock returns can appear over time. Moreover, the market cannot be perfectly efficient or there would be no incentive for professional investors to uncover the information that gets so quickly reflected in market prices. (Ibid.)

1.2 Regression to the mean phenomenon

Regression to the mean, sometimes called mean reversion, is a statistical phenomenon that can make natural variation in repeated data look like change. It happens when unusually large or small measurements tend to be followed by measurements that are closer to the mean. The phenomenon was discovered and named late in nineteenth century by Sir Francis Galton. He published an article in 1886 under the title “Regression towards Mediocrity in Hereditary Stature”. The article reported measurements of size in successive generations and comparison of the height of children to the height of their parents. The experiment showed that the offspring did not tend to resemble their parents size, but to be always more mediocre than they; to be shorter than the parents, if the parents were unusually tall; to be taller than the parents, if the parents were very short. (Kahneman 2011, 179)

Regression to the mean phenomenon is widely recognized in hypothesis testing, validity and threads of validity. The concept of regression to the mean suggests that whenever participants are selected because their scores on a measure are either very high or very low, they will tend to be less extreme on a second testing; that is, their scores will have regressed toward the mean (Graziano and Raulin 2004). How much regression occurs will depend on how much of the test performance is due to variable factors and how much is due to consistent factors such as skills and learned habits. Kahneman states that success is a combination of talent and luck (2011, 177). Luck can be seen as another meaning for variable factor. The more these variable factors contribute to the score, the more regression you can expect to see.

It is very unlikely to observe data without regression to the mean phenomenon but the effect of regression to the mean can be reduced by a good study design. Barnet et al. present two such designs: random allocation to comparison groups and selection of subjects based on multiple measurements. If participants were randomly distributed to comparison groups, the answers from all the groups should be equally affected by regression to the mean. With placebo and treatment group the mean change in the placebo group provides an estimate of the change caused by regression to the mean. The difference between the mean change in the treatment group and the mean change in the placebo group is then the estimate of the treatment effect after adjusting for regression to the mean. The second study design method to reduce regression to the mean is selection of subjects based on multiple measurements. The effect of regression to the mean increases with larger measurement diversity. To reduce the diversity participants can be selected using two or more baseline measurements. This method can be thought of as an attempt to get a better estimate of each participant´s true mean. (2005, 217-218)

Murstein has studied how regression to the mean phenomenon affects stock market. In his article, Regression to the mean: One of the Most Neglected but Important Concepts in the Stock Market, he demonstrated the effect of regression to the mean in stock market with two previous studies. Table 1 shows how various Morningstar mutual funds performed in 1989 and in 1994. It became evident that all the returns of the funds above the mean in 1989 were below the mean in 1994 and vice versa. The fact that such an unusual switch occurred implies a lot of random error affecting fund performance, and probably only a small amount of true ability involved. (2003, 235)

Table 1. Comparison of returns of mutual funds by classification (percentage)

Objective	Five Years to March 1989	Five Years to March 1994
International Stocks	20.6	9.4
Income	14.3	11.2
Growth and Income	14.2	11.9
Growth	13.3	13.9
Small Company	10.3	15.9
Aggressive Growth	8.9	16.1
Average	13.6	13.1

Source: (Murstein, B. The journal of Behavioral Finance, 2003 Vol. 4, No. 4, p 235)

To demonstrate that the above case is not atypical, Murstein gives another example of the effect of regression to the mean in stock market. A sample of 283 equity funds were studied from year 1990 to 1994 and identified the twenty funds with the highest returns. These twenty funds outperformed the S&P 500 index by an annual average of 9.2%. The returns of these stocks were recalculated for years 1995-1999. The result showed that these formerly best performed funds trailed the S&P 500 index by an annual average of more than 2%. Table 2 shows how the top ten funds of the 1990-1994 period ranked over the 1995-1999 period. The rank of these funds is between places 21 to 275. (2003, 235)

Table 2. A shift in performance: early 1990s versus late 1990s

Fund	Rank in 1990-1994	Rank in 1995-1999
A	1	129
B	2	134
C	3	261
D	4	21
E	5	210
F	6	53
G	7	183
H	8	105
I	9	275
J	10	54

Source: (Murstein, B. The journal of Behavioral Finance, 2003 Vol. 4, No. 4, p 235)

In both Murstein´s examples, the volatility of the stocks was quite high and stocks may even have switched sides from the best performing to the worst performing position. Radical changes like these are not always typical. Pzena demonstrated regression to the mean phenomenon by making a sample of the S&P 500 index for 20 years' period. The data was divided into five quintiles based on each company´s return on equity (ROE). The lines show the quintiles´ performance during the following five years (Figure 1). The Q1 line represents the least profitable companies in the base year and traces their succeeding five-year performance. In the base year, the Q1 companies are basically making no money and zero return on equity. After the five-year period, they have 10% ROE. Of course, some companies do not see the brighter future and end up bankrupt. The lines show the average for each quintile. On the upper side of the lines, is the Q5 which presents the most profitable companies. Their average ROE was 25% in the base year. After the five years, the return was 15%. Even though, regression to the mean can be detected in the Figure 1, the lines hold their positions trough the five-year period. The quintile of the best performing companies maintained its position as did the worst performing quintile. (1995, 38)

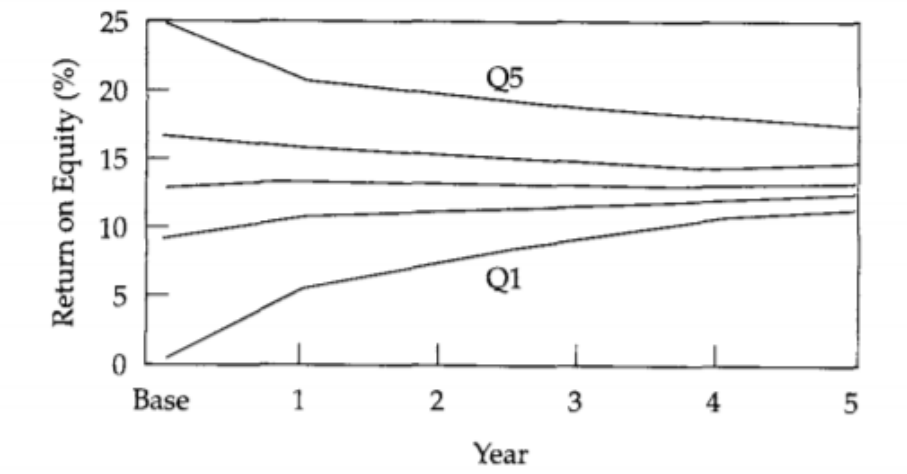

Figure 1. Regression to the mean by the S&P 500 index quintiles

Source: (Pzena 1995, 38)

Kim, Nelson and Startz (1991) examined the persistence of mean regression in stock returns over the pre-war and post-war (after the Second World War) periods in the United States. They found evidence of mean regression during the pre-war period, but did not find evidence of mean regression during the post-war period. Based on this empirical evidence, they concluded that mean regression in the United States stock market is primary a pre-war phenomenon. In contrast, when Poterba and Summers (1988) examined the behavior of stock returns in 15 countries using primarily post-war data (1957-1986), they found that 13 out of 15 foreign stock markets exhibit mean regression over long horizons. Sauer and Chen implemented similar test than Kim, Nelson and Startz but in the United Kingdom. They examined the behavior of a total return stock market index for the United Kingdom over the period 1919 through 1990. They found no evidence of statistically significant mean regression when the entire time series was examined. However, again when the returns were separated into pre-war and post-war periods, they found evidence of statistically significant mean regression during the pre-war period. In the absence of transaction costs, it may have been possible to gain abnormal returns during the pre-war period. However, the empirical evidence suggest that an investor cannot use the information contained in the pre-war period performance to consistently earn abnormal returns. Therefore, Sauer and Chen interpreted

their empirical evidence as consistent with the efficient market theory. (Sauer and Chen 1996, 1381)

It has been studied if regression to the mean occurs in the Standard & Poor´s 500 index basis changes, and can it be used as an arbitrage opportunity or is it statistical illusion. Mean regression in stock index basis changes has been amply documented by several professors and researchers. Miller, Muthuswamy and Whaley propose an alternative explanation for the observed negative correlation in basis changes that it is mainly a statistical illusion than arbitrage because many stocks in the index portfolio trade infrequently. (1994, 506)

According to Samsa there are two schools of thought: nihilists and fundamentalists. The claim that all available information is contained in the current stock price is essentially nihilistic. As it claims that the fundamental elements of the company (e.g., its business model and management) do not matter. Fundamentalists argue the opposite: not only does this information matter but, although it might be ignored by the market in the short-term, truth about long-term corporate performance will eventually win out. There is one point about which these two schools of thought agree. The stocks of safe-haven companies have lower risk than average. The nihilists base their opinion on the low level of volatility in their stock prices. Fundamentalists base their opinion on the fact that these companies are large, well-financed and offer products that consumers continually purchase. The Coca-Cola Company, Johnson & Johnson and Consolidated Edison are examples of such safe-haven copmanies. In practice, the prices of safe-haven companies evidence a stronger tendency to regress to the mean than do most other companies. One explanation for this phenomenon is that, because their stream of earnings and dividends is so predictable, their fundamental value can be estimated with mush higher degree of accuracy than it is the case for many other companies. Indeed, the prices of these stocks most often depart from their fundamental value because of the view that large institutional investors have about the other stocks in the market. (2014, 107)

The goal of safe-haven investors is to buy safe-haven companies at price below their fundamental value, to hold these companies until their prices exceed their fundamental values, and thus to obtain superior risk-adjusted returns. Nihilists deny that it would be possible for any kind of strategy to obtain superior risk-adjusted returns. In general, and consistent with the broader literature of behavioral finance, such value-based strategies tend to outperform the market, but only

a modest amount. The nihilists reply that the modest levels of outperformance are evidence of how difficult it is to beat the market. (Samsa 2014, 107)

There are countless examples of regression to the mean phenomenon and how it can be detected in studies such as the comparison of the height of children to the height of their parents or in the stock market. According to Pzena the reason why regression to the mean occurs is human nature. In the S&P 500 index quintile example the managers and owners of the best performing companies and the worst performing companies are dealing with very different problems. The managers of the best performing companies must deal with the competitors that want to replicate the successful strategy and get what the company has: customers, product or reputation. No matter how hard the managers resist, history has showed that those competitors will get a piece of it. The successful company´s growth rate will slow, and its profit margins will fall. On the other hand, the managers of the worst performing companies are usually working hard to raise their companies from the bottom, and make them successful. The ways to do so vary from introducing new products to making changes in the management team. (1995, 38)

Barnet et al. explain the regression to the mean: "It happens because values are observed with random error. By random error we mean a non-systematic variation in the observed values around a true mean... Systematic error, where the observed values are consistently biased, is not the cause of RTM. It is rare to observe data without random error, which makes RTM a common phenomenon." (2005, 215) Random error is an error in measurement that cannot be predicted and the causes may not be identified. Sauer and Chen present two economic models that have been put forth to explain the existence of mean regression in the stock market returns. One argument suggests that mean regression reflects the impact of rational investors time-varying expected returns. This interpretation of mean regression is consistent with efficient market. Another argument suggests that mean regression in stock returns results from an inefficient market where stock prices temporarily diverge from their underlying fundamental values. (1996, 1392-1393)

Many scholars have their own opinions and arguments regarding the mean regression as a phenomenon and its appearance in the stock markets as well as the reason behind the phenomenon. Others claim that it is only a pre-war phenomenon when others claim that it is just a statistical illusion. There is not a uniform argument about the regression to the mean phenomenon which makes it an interesting topic to research.

1.3 Stock market

The stock market, also known as the equity market, is the market in which shares of publicly held companies are issued, bought and sold. Stock market can be divided into two parts. The primary market is a stock market where companies issue new shares to the public in an initial public offering to raise capital. Once these new shares have been sold in the primary market, they can be traded in the secondary market. There investors buy shares from other investors at the market price or at any price the both participants agree upon. The stock market is one of the most important ways for companies to raise capital. For exchange investors get part of the company´s ownership for themselves. The total value of world stock market is tens of trillions of dollars. The amount has increased steadily over the last years. The biggest stock market exists in the United States of America followed by United Kingdom and Japan when measured in market capitalization value (Worldatlas.com 2016).

There has been predecessors and examples of stock markets as early as in the 12th century. In Italy bankers began to trade in government securities, and in France the first brokers began to manage and regulate the debts of agricultural communities on behalf of banks. The first genuine stock market was established in Belgium a few hundred years later. However, these early stock markets did not trade stocks but instead the markets dealt with the affairs of government, business and individual debts. The system and organization were similar but the actual properties traded were different. The world´s first publicly traded company was the Dutch East India Company in the beginning of 16th century. The company was focused on trade, exploration and colonization. The company and its easily transferred shares created Amsterdam stock exchange, the first stock exchange in the world.

There are 16 stock exchanges whose market capitalization is more than one trillion dollars. The most noteworthy ones are New York Stock Exchange (NYSE), NASDAQ and London Stock Exchange. Some exchanges, such as NYSE and London Stock Exchange, have a physical location in which trades are made but other exchanges are existing only electronically, an example would be NASDAQ. A stock exchange allows companies to raise funds by connecting them with private and institutional investors. Stock exchange has many roles including bringing companies and investors together, enabling issuers and companies to raise new capital, organizing and overseeing

a fair and efficient market, and providing timely and accurate trading and company disclosure information to inform private investor trading. (London Stock Exchange 2014)

The development of stock markets is affected by countries´ economic development, political decisions, natural disaster and conflicts between people, just a few to mention. When discussing the stock indices, the long-term trend has been upward but short-term trend has been random walk. The Figure 2 illustrates the total returns of stocks, long- and short-tern government bonds, gold and commodities between years 1802 and 2003. One dollar invested in stocks in year 1802 would have grown to $8.8 million in 2003 if all dividends had been reinvested. The average compound after-inflation rate of return on stocks from 1802 to 2002 was 6.8% per year, and the number has remained remarkable steady over time. This return exceeds returns of all other financial assets. Historically investment in stocks has been more profitable compared to other assets. Based on the data from the last two centuries it could be assumed that investment to stocks would be profitable also in the future. However, the past performance does not guarantee future performance. (Siegel 2008)

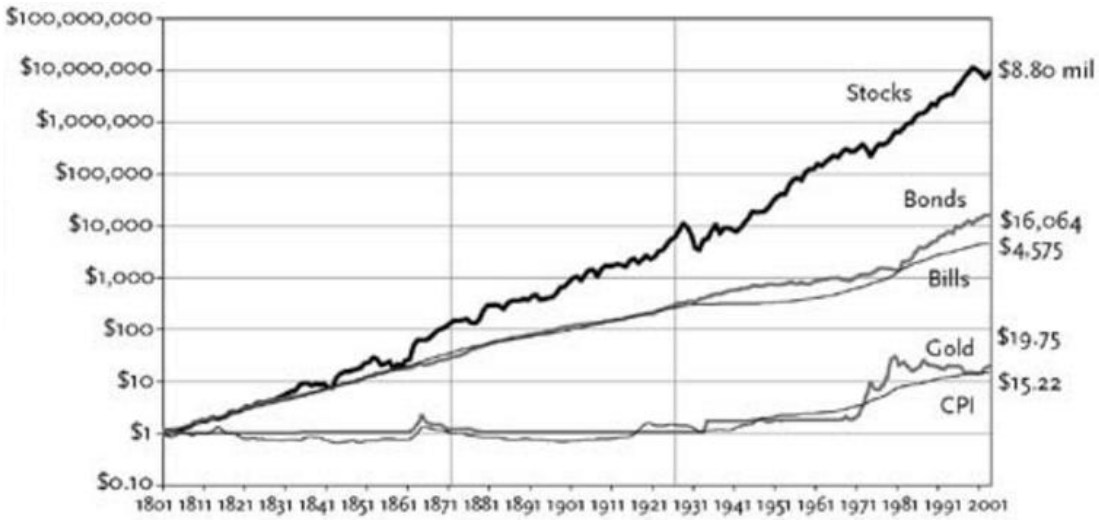

Figure 2: Total nominal return indices 1802-2003

Source: (http://www.econlib.org/library/Enc/StockMarket.html)

1.4 The Standard & Poor 500 index

Standard & Poor´s Financial Services LLC is an American financial services company which was founded in 1860 by Henry Varnum Poor. In 1941 Poor´s Publishing company merged with Standard Statistics, and adopted the name Standards and Poor`s Corporation. The company is based in New York with offices worldwide, and it operates as a subsidiary of S&P Global. The company provides market intelligence in the form of credit ratings, indices, research and thorough leadership for investors worldwide. Its services are consumed by corporates, financial institutions, government, insurance, structured finance, energy, media and communication, metals and mining, and real estate sectors. (Bloomberg 2017)

All the S&P U.S. indices are designed to reflect the U.S. markets and the U.S. economy. The S&P 500 index consists of 500 companies whom stocks are traded on New York Stock Exchange (NYSE), American Stock Exchange (AMEX) and the Nasdaq National Market System (NASDAQ). Although the index contains 500 companies, it may contain greater than 500 stocks since some companies may be presented by multiple share classes in the index. The index was launched in 1957. The S&P 500 focuses on the large-cap sector of the market. Companies in the S&P 500 are considered leading companies in leading industries. (S&P Dow Jones Indices 2017a, p. 4). When dividing the index into sectors, the biggest sector with a 21.5% share is information technology. For example, Facebook Inc., Apple Inc. and Microsoft Corp belong to this sector. The second biggest sector is financials (14.8%) including Berkshire Hathaway, JP Morgan Chase & Co, just a few to mention. Other sectors represent health care, consumer discretionary and industrials among others. (S&P Dow Jones Indices 2017b)

Figure 3 demonstrates the S&P 500 index between years 1977 and 2017. There have been two major worldwide incidents that have had remarkable impact on the index. First, there was the Dotcom Bubble in 1995-2000 which was characterized by a rapid rise in equity markets fueled by investments in internet-based companies. The bubble burst in 2000. (Geier 2015) After the burst another boom began. The S&P 500 index reached its record high point so far in 2007, and almost immediately after that the worldwide financial crisis began. It took seven years to recover to the level where the index was before the financial crisis. (Havemann 2009) Since then has been ascending and it has reached its highest point of all times in 2017.

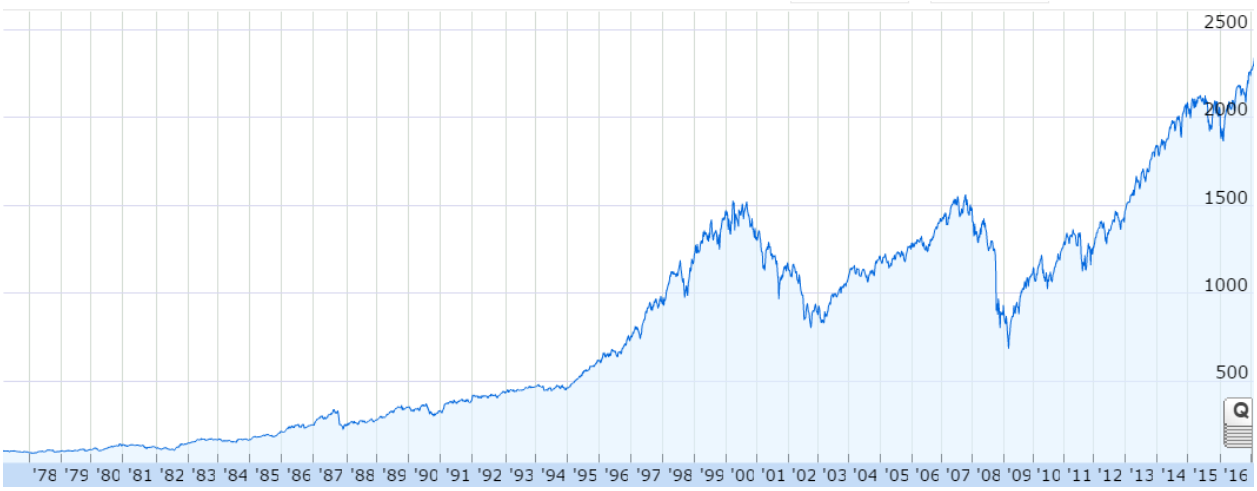

Figure 3. S&P 500 index 1977-2017

Source: (Google Finance)

No company can apply or be nominated for inclusion into the index. The selection process is performed by an autonomous S&P Index Committee, a team of analysts and economists. Changes to the S&P 500 index are made as needed, with no annual or semi-annual reconstruction. The list below presents eight criteria that a company must fulfill to be included to the S&P 500 index.

Eligibility criteria (S&P Dow Jones Indices 2017a, 5-6):

1) Market Capitalization - Unadjusted company market capitalization of US$ 5.3 billion or more.
2) Liquidity - Adequate liquidity and reasonable price.
3) Domicile - Only U.S companies can be included.
4) Public float - Public float at least 50% of the stock.
5) Sector classification - Sector balance for the index must be maintained after inclusion.
6) Financial viability - The sum of the most recent four consecutive quarters´ Generally Accepted Accounting Principles (GAAP) earnings should be positive.
7) Treatment of IPOs - Initial public offerings should be seasoned for 6 to 12 months before being considered for addition to an index.

8) Eligible securities - The company must not be a closed-end fund, holding company, partnership, investment vehicle or royalty trust.

Companies which have been added to the index usually stays in the index. Standard and Poor has stated that the turnover in index membership should be avoided whenever possible. Since the foundation of the index there has been more than 1000 stocks or companies that have been replaced by other stocks or companies in the S&P 500 index. There are two main reasons why companies have been deleted from the index. The first reason is if the company substantially violates one or more of the eligibility criteria. The second reason why companies are deleted from the index is that due to the mergers, acquisitions or significant restructuring companies no longer meet the inclusion criteria. When the company is removed from the index, S&P prepares an explanation of the removal. (S&P Dow Jones Indices 2017a, 6-7)

2. RESEARCH ON REGRESSION TO THE MEAN IN THE STOCKS OF THE S&P 500 INDEX

Based on the literature study, the author has come to conclusion that regression to the mean phenomenon can be detected in the stock market. How strongly it exists, is widely discussed. Under some circumstances the regression to the mean has been detected whereas in most cases the degree of mean regression is insignificant i.e. there is no obvious arbitrage opportunity available. In this chapter the regression to the mean of the stocks in the S&P 500 index is tested. The S&P 500 index was chosen to be the base of the study because it is one of the best-known indices in the world. It contains 500 companies and their stocks which makes it an excellent reflector of the U.S. economy and the U.S. market. The index has been used in studies before which gives a good base for the comparison and adds to the current literature as it includes latest crisis episode. The author hopes to bring clarification to the problem, is there regression to the mean in the stocks of the S&P 500 index or not. The author aims to clarify if regression to the mean can be detected in the in the stocks of the S&P 500 index. The purpose is to gather a large sample which is then tested to highlight regression to the mean in the index. The main hypothesis is that regression to the mean is detected in the stocks of the index.

Secondly, if regression to the mean is detected in the stocks of the S&P 500 index, the author will study is there arbitrage opportunity available. As there is a debate if regression to the mean phenomenon exists in the stock market, there is also discussion on how relevant the regression is. Some previous studies have showed significant regression to the mean in the stock market. On the other hand, the detected regression can be so minor that taking transaction costs and taxes into consideration, the overall return may be same or similar to the total S&P 500 index´s return. In that case, there would not be incentives for investors to use regression to the mean

phenomenon to gain profit. Based on the results of the study it can be assessed if regression to the mean is as significant that it can be used as arbitrage opportunity. The author does not take a side on how much an investor can gain abnormal returns calculated in dollars. Each investors´ transaction costs and taxes have an impact on overall return but those factors are not calculated in this thesis because amount of transaction costs and taxes are unique factors of each investor.

Not all tools are available for the author to investigate the regression to the mean phenomenon. Some useful tools and programs are not free of charge and require to be bought with a license. This kind of tools can be quite expensive that is why the author decided to execute the research with programs and resources that are already available for her to use and which the author is familiar with. For example, Stata which is integrated statistical software for data analysis, data management and graphics is one of the tools which would have been an excellent tool for analyzing mean regression. Unfortunately, it would have cost almost two hundred euros for the author to purchase the software. Although, some programs would have been chargeable to the author, but she was able to use them since those were offered by the university. The author recognizes that research may have been easier, more thoroughly and broader investigated if more professional tools were available.

2.1 Research strategy

Kothari defines two main approaches to research: quantitative approach and qualitative approach. The third option is the mix of quantitative and qualitative approaches. Quantitative approach involves the generation of data in quantitative form which can be subjected to rigorous quantitative analysis in a formal and rigid fashion. Quantitative research focuses on gathering numerical data and generalizing it across groups of people or to explain a certain phenomenon. This approach can be further sub-classified into inferential, experimental and simulation approaches. The purpose of inferential approach is to form a database from which to infer characteristics or conclusions of population. Experimental approach is characterized by greater control over the research environment and in this case, some variables are manipulated to observe

their effect on other variables. Simulation approach involves the construction of an artificial environment within which relevant information and data can be generated. (2004, 5)

Quantitative approach is chosen to identify if regression to the mean phenomenon is detected in the sample of the S&P 500 index. The mean regression is a numerical phenomenon, and that is the reason why quantitative approach suits best for this thesis research. The research is implement by analyzing a database of numerical data to define if regression to the mean phenomenon is detected. From the sub-classes of the quantitative approach the inferential approach is the one that best describes the purpose of the thesis research. The sample of population is studied to determine its characteristics, and it is then inferred that the population has the same characteristics.

The relevance of hypotheses to the study is the main distinctive point between research approaches claim Bryman and Bell. They divide the research approaches into three types: deductive, inductive and abductive approach. The author finds that the deductive approach is the most suitable for this research. The deductive approach tests the validity of assumptions, theories and hypotheses. The research does not try to create new theories or overall stock market generalizations based on the research results. The research aims to confirm or reject the research hypotheses and is then following the deductive approach. The thesis conforms the following structure: theory, hypotheses, data collection, observation/test and confirmation/rejection. This is a common structure for deductive research approach. (2015, 23-24)

2.2 Research design

"A research design provides a framework for the collection and analysis of data" (Bryman and Bell 2015, 49). "A good research design often possesses the qualities such as being flexible, suitable, efficient, economical. Generally, a research design which minimizes bias and maximizes the reliability of the data collected and analyzed is considered a good design." (KL University, 311) The author believes that descriptive statistical research design best fits to solve the research questions because a study that is concerned with specific predictions or with the narration of facts and characteristics relating to an individual, group or situation, are examples of descriptive research

studies. (KL University, 309) Descriptive research attempts to determine, describe or identify rather than explaining causes and effects. The author tries to solve if regression to the mean phenomenon is detected in the S&P 500 index. The author does not try to determine what causes the phenomenon. On the other hand, the author will point out her opinions to the question if regression to the mean can be used as an advantage to gain abnormal returns or not. For the question, how much abnormal returns it is possible to gain because of the phenomenon, the author does not answer. Therefore, detailed explanation of the effects of the phenomenon are not discussed in the thesis.

By the decided descriptive statistical research design, the author finds that quantitative research method best suits to clarify the research problem. "Quantitative methods emphasize objective measurements and the statistical, mathematical or numerical analysis of data collected through polls, questionnaires and surveys, or by manipulating pre-existing data. Quantitative research focuses on gathering numerical data and generalizing it across groups of people or to explain a particular phenomenon." (University of Southern California 2017) The research database consists of quantitative information of the companies included in the S&P 500 index. The existing data is processed to define the existence of regression to the mean phenomenon on the index.

2.3 Data collection, database and time period

The data collection of this research consists of two main sources: primary and secondary data sources. The primary sources allow researchers to get as close as possible to the original ideas, events, and empirical research as possible. Secondary sources analyze, review, or summarize information in primary resources or other secondary sources. (Virginia Polytechnic Institute and State University 2017)

For the research, the author used secondary source to collect the data. The data was collected through Thomson ONE program and Microsoft Excel add-in provided by Thomson Reuters. The data was collected by creating a formula in the Thomson ONE excel add-in to produce a list of stocks and companies in the S&P 500 index. The list of stocks includes companies that

were in the S&P 500 index on 29.03.2017. With the program, the author gathered main key ratios from each company to investigate the regression to the mean phenomenon.

During the data collection, the author had to make adjustments regarding the desired data due to the limitations of author´s capabilities and limitations of available data sources. More thorough description is on the chapter Limitations of the research.

For the database, the author chose the S&P 500 index in the Standard and Poor´s index family. The index contains 500 leading companies and their stocks, and captures approximately 80% coverage of available market capitalization. The index is widely regarded as the best single gauge of large-cap U.S equities. (S&P Dow Jones Indices, 2017b) It is designed to reflect the U.S. market and the U.S. economy. The index expanded to its current form (containing 500 companies) in 1957 therefore, it has a long history as reflecting the stage of the U.S economy. The S&P 500 index has been used as a database in previous studies as well regarding the regression to the mean phenomenon. The results have been both for and against the phenomenon. The author chose the index since it contains high amount of companies from different fields of business. Secondly, versatile data is available about the index and the companies, because of the long existence of the index.

For the time frame, the author chose data between years 2000 and 2015. The author felt that the 15-year time frame was long enough to get relevant monthly and yearly data and make conclusions about the regression to the mean phenomenon. In 2000, the index reached its all-time high point just before the Internet bubble burst. During the next couple of years, the index declined but in 2002 the new boom began. After that the worldwide financial crisis hit the U.S economy. After the recovery in 2009, the index has continued its rise and has reached its record high point in its whole history. There have been many phases in the development of the index in the past 15-year time frame. The author believed that it would generate versatile data for the research.

2.4 Sampling design and description

The sample size of 505 stocks is gathered from a finance database through the Thomson ONE program, a trading and investing tool which delivers a broad range of financial content such as company and industry information, market news and reference data. (Thomson Reuters) The author chose this platform for gathering her sample because it is one of the most comprehensive and versatile platform to research company and market information. The author also found it easy to use with the excel add-in that was provided by the Thomson Reuters to gather and analyze data in excel format.

The author created a list of stocks and companies that were in the S&P 500 index on 29.03.2017. Based on the list provided by the Thomson ONE, there were 505 stocks in the S&P 500 index. With the names of the stocks and companies, the program also provided ticker symbols of each stock. A ticker symbol is an abbreviation used to identify publicly traded shares of a particular stock exchange. There were multiple ready-to-use formulas in the program, and with the ticker symbol it was possible to calculate any key ratio needed.

For the research, the author decided to use the stocks in the S&P 500 index because the companies, of which stocks can be found in the index, represent multiple business sectors which makes the sample versatile. Financial, information technology and health care sectors are the three biggest sectors represented in the index. The author chose to create a list of stocks and companies in the S&P 500 index on 29.03.2017 because the only publicly available information of the content of the S&P 500 index that the author found, is the content at the current moment. For more thorough and detailed research about regression to the mean phenomenon, it would be useful to gather the content of the index of each period investigated.

In Appendix 1, the whole list of stocks and companies in the S&P 500 index on the 29.03.2017 is presented with separate columns showing each company´s name and stock ticker number.

2.5 Analysis of the mean regression

The research was designed by the author for the most accurate complementation of the aim of the research. The quantitative research focuses on gathering numerical data to explain a regression to the mean phenomenon. The author assembled a database of companies, stocks and those total returns from period 2000 to 2015. The database was then examined in Microsoft Excel. The research has all together four parts, all of which composed to give a relevant and distinct viewpoint.

The author aimed to make the research as precise as possible to give an opinion if regression to the mean can be detected in the stocks in the S&P 500 index or not. Previous studies have showed that regression to the mean phenomenon has been detected in the stock market. Other researchers claim in their studies that the mean regression can be detected but it is too negligible to be used as an advantage to gain abnormal returns. Finally, some researchers say that markets are efficient and regression to the mean is not detected in the stock market. To get the most precise interpretation of the existence of the phenomenon, the author focused on four main areas in the research:

1) Average yearly total returns
2) Average monthly total returns
3) Decile movement analysis
4) Regression analysis

The author designed the content of the research in a way that it would best answer to the research question of the thesis; is regression to the mean phenomenon detected in the stocks in the S&P 500 index. The first part is in place to reveal the yearly total returns of the stocks. The time frame covers years between 2000 and 2015. Each year, the stocks are divided into ten equally sized deciles based on stocks´ total return of that year. The rank is done from best performing stock to worst performing stock. The average total return of each decile is calculated, and summarized into a table with a five-year period (e.g. base year 2000 and years 2001-2005 as a study period). Then, the test is repeated with the following years. After the procedure, the author had eleven tables assembled as mentioned. Therefore, the average total returns of each decile shown in the tables of

the following chapters is an average of eleven averages. A line chart is created to demonstrate the movement of average total returns of each decile, and to illustrate the possible mean regression. The author believes that a five-year period is significant enough to detects the mean regression. No longer period is needed.

The second part is constructed similar as the first part with a difference that, instead of yearly total returns, monthly total returns are used in this section. The time frame is twelve months (one year). The stocks are again divided into ten deciles based on stocks´ monthly total returns from best performing to worst performing. Again, a summary table of each deciles average total returns and figure is created to illustrate the mean regression. The average total returns are calculated with year and month precision to get two different viewpoints to investigate if mean regression is detected in the stocks in the S&P 500 index. The monthly research is more specific than a yearly research, and it was executed in addition the yearly research to have two different time frames.

The third part aims to clarify the summary tables of the parts one and two, and to illustrate the movement of the deciles. In the summary table of yearly average total returns, the ten deciles were divided based on the average total returns of the base year, T0. The table in part three, illustrates the decile rank in year T1 when the average total return deviation was done based on the performance in year T0, and what is the rank of year´s T0 deciles in year T1. For example, what is the rank of no. 1 decile of a year T0 in a year T1. The rank is repeated with years T2, T3, T4 and T5. The same is done with the table of monthly average total returns. The base month is T0 and the rank is done with months from T1 to T11.

The fourth and last part examines the possible mean regression with regression analysis tool available in Microsoft Excel. The author decided to use this analysis as a final tool to reveal if regression to the mean exists in the average total returns of stocks in the S&P 500 index. With this tool, the author believes that it is possible to find the smallest hint of mean regression which would not be possible to detect with the previous studies in part one, two and three. This is the most accurate tool of all methods used by the author. The regression analysis tool will provide statistical and numerical data of possible mean regression. The regression analysis tool is used with both yearly and monthly data.

The research was designed from a broadest period (yearly) to the narrower period (monthly). The author believes that this is the right practice to proceed with the research. The last part, the regression analysis tool, will give the most accurate data of mean regression. With these four tools, the author hopes to identify if mean regression can be detected and if yes, how significant the regression is.

2.6 Data analysis methods

The sample is being statistically analyzed by the author by utilizing Microsoft Excel. The data analysis of this research is approached by focusing on one part at the time which will give a clear framework for the analysis process. Accordingly, the data analysis process is divided into four main parts which were presented in the previous chapter. The data analysis approach is quantitative and two data analysis methods are used: descriptive statistics and linear regression. A graphical presentation will be provided, and each research part is separately interpreted with the aim to show possible mean regression in the data. The data analyzing will contain both linear and scatter plot figures.

The main focus of the research and data analysis is to identify the mean regression by analyzing and comparing the means of each decile´s total returns. Mean is one of the main measures of central tendency which is used in descriptive statistics method. This method is used in the first three parts of the research. The data analysis of the first two parts will contain tables of data, and line chart figures to illustrate the average total returns. In the third part the descriptive statistics will concentrate on the movement of the ranking of total returns of each decile.

The second method used in the research is linear regression. This method is used in the final part of the research. With the regression analysis tool, the author gets the most accurate information about the relationship between the variables. The regression analysis tool will provide information of how much a performance of one period affect to the performance of the following period. After implementing the regression formula, the author will have numerical data about the significance and validity of the regression. The results are presented with illustrative scatter plot figures. The author decided to use 95% confidence level. The confidence level states that if the research was

repeated, the results would match the results from the actual population 95% of the time. The author chose the 95% confidence level because it is one of the most used confidence level, and it is accurate enough to give statistically reliable results.

2.7 Limitations of the research

The original idea was to collect information of the companies in the S&P 500 index in each specific year so that each year, between 2000 and 2015, the data would have contained all the companies in the index in that specific year. During the data collection process, the author was unable to find this kind of information from the sources available to her. The portfolio of companies from year 2000 to 2015 is composed based on the company listing in the S&P 500 index on 29.03.2017. When making the list of companies in a specific year, the companies that did not exist in the index on that time were deleted from the portfolio. Due to this procedure, the amount of companies varies between 410 and 505 each year. Due to this the portfolio of companies in each year and the average total returns are not comparable with the index return of that year.

3. RESEARCH RESULTS AND DISCUSSION

The third and final chapter of the thesis will cover the analysis and discuss the findings of the study as described in the second chapter. As mentioned, the findings will be presented in four separate paragraphs. Each section will emphasize the most important findings contributing to the main research agenda, which is to find out if regression to the mean phenomenon can be detected in the stocks of the S&P 500 index. After presenting the results, the author will discuss about the results and give suggestions for possible following research of the regression to the mean phenomenon.

3.1 Average yearly total returns

The first part in investigating the possible mean regression in the stocks of the S&P 500 index was to examine average yearly total returns. This part aimed to find out how the average total returns have changed over five-year period when decile distribution was used based on the performance of the base year, T0. The general index return of each year had been reduced from the stocks and deciles´ total returns. By doing this way, only the companies´ own performance affected to the total return of the stocks. The distribution had been made from best performing decile to the worst performing decile in year T0 when the average total return varied between +85.68% and -50.87%. The table 3 summarized the average yearly total returns in five year-period.

Table 3. Average yearly total returns

Decile	T0	T1	T2	T3	T4	T5
1	85,68 %	-0,75 %	8,84 %	5,22 %	2,58 %	2,78 %
2	29,25 %	-4,51 %	-0,96 %	-0,85 %	0,15 %	-1,71 %
3	14,62 %	-2,69 %	-2,28 %	-1,65 %	-1,55 %	-0,52 %
4	5,10 %	-1,97 %	1,03 %	-0,95 %	0,96 %	-0,56 %
5	-2,34 %	-1,89 %	-4,30 %	-3,06 %	-1,74 %	-3,96 %
6	-9,26 %	-1,31 %	-2,80 %	-0,89 %	-0,90 %	-3,04 %
7	-16,23 %	-0,44 %	-2,64 %	-1,73 %	-2,14 %	-0,50 %
8	-23,42 %	-2,59 %	-2,22 %	-2,70 %	-0,43 %	-0,07 %
9	-32,52 %	1,88 %	1,40 %	-0,29 %	-0,96 %	2,32 %
10	-50,87 %	14,27 %	3,94 %	6,90 %	4,02 %	5,26 %

Source: Composed by the author

As can be seen from the table above, the distribution of average total returns between the deciles has been narrowed significantly compared to the distribution of the base year. In year T1 the distribution of average total returns varies between +14.27% and -4.51%. In year T2 the average total returns vary between +8.84% and -4.30%. In year T3 the difference of average total returns between the best performing decile and the worst performing decile is 9.96 percentage points. In year T4 the difference is at its narrowest level, only 6.16 percentage points. In the final year of the five-year period, the distribution increases a bit. During the five-year period, none of the deciles of any particular year reach as high or low average total returns as did deciles in the base year. As a conclusion, the average total returns and the difference between them are much smaller each year during the whole five-year period than in the base year.

The mean regression is clearly detected in the change of the average total returns between T0 and T1. The average total returns of the four best performing deciles in T0, decline compared to the next period, T1. At the same time, the average total returns of the six remaining deciles increase in the period T1. These two periods are the only periods where a clear regression to the mean is detected through the whole sample of the ten deciles. During the following periods, the change of direction of the average total returns has been random walk. For example, the average total return of decile 10 is -50.87% in the base year. Next period, the average total return increases to +14.27%, and next period it declines again to +3.94%. The whole five-year period has been a

variation of increase and decline of the average total returns. Therefore, the decile has not had one clear direction.

The following Figure 3 presents the same average yearly total returns as shown in the Table 3 but in a more illustrative form.

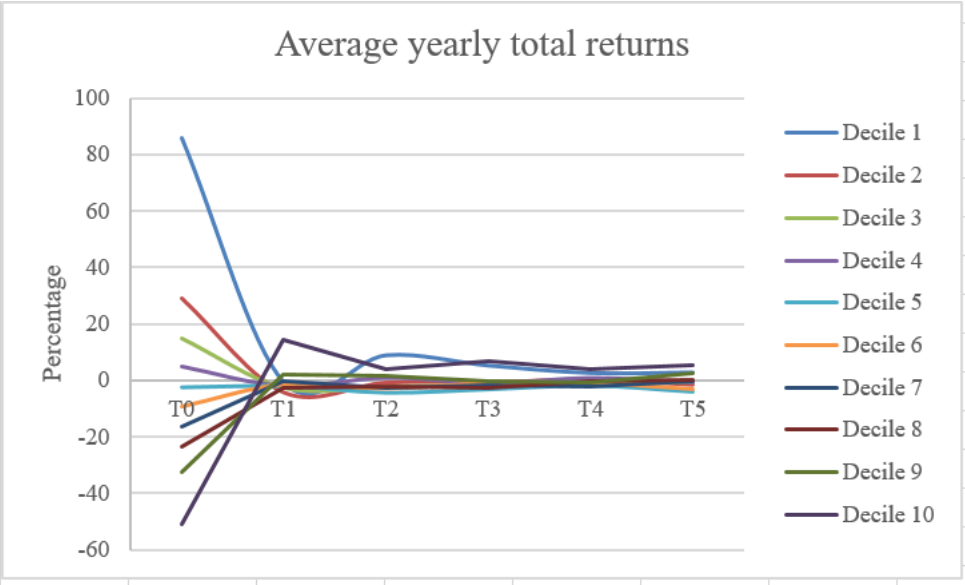

Figure 3. Average yearly total returns presented in the line chart

Source: Composed by the author

The Figure 3 shows how clearly the mean regression has been detected in the period T1. None of the deciles continues performing at the level where they are at the base year. The performance of decile 5 has been almost horizontal during the whole examination period, the average total return has stayed close to zero percent. All other deciles (excluding deciles 1 and 10) have had a minor variation on the both sides of the zero level. The performance of deciles 1 and 10 is abnormal compared to the performance of other deciles. The decile 1 has a significant decline in average total return when moving from T0 to T1. After this, the performance has not been as impressive as it was at the beginning but the decile stands out from other deciles with its performance. The decile 10 has the best average total return in year T1, and during rest of the examination period it holds its position being as one of the best performing deciles.

3.2 Average monthly total returns

The second part is similar to the previous chapter but now the focus is on average monthly total returns instead of yearly. By having two different periods, the author gets more detailed and specific information about the possible mean regression. The general index return of each month has been reduced from the stocks and deciles´ total returns. By doing this way, only the companies´ own performance affects to the total return of the stocks. The period covers the base month (T0) and following eleven months (T1-T11), altogether the period is one year. The base month is January and the table ends to December. All other months are between them in their calendar order. Again, the distribution has been made from best performing decile to the worst performing decile in in base month T0 when the average total return varied between +20.21% and -16.28%. The Table 4 summarized the average monthly total returns in eleven month-period.

Table 4. Average monthly total returns

Decile	T0	T1	T2	T3	T4	T5	T6	T7	T8	T9	T10	T11
1	20,21 %	0,09 %	-0,07 %	0,59 %	-0,71 %	1,84 %	-0,63 %	0,40 %	0,21 %	1,68 %	-0,17 %	0,27 %
2	8,93 %	-0,91 %	-0,04 %	-1,18 %	-0,68 %	-0,46 %	0,18 %	-0,61 %	-0,17 %	0,73 %	0,54 %	-0,38 %
3	5,03 %	-0,45 %	-0,98 %	-0,92 %	0,39 %	0,61 %	0,06 %	-0,37 %	0,22 %	-0,49 %	-0,28 %	0,40 %
4	2,45 %	1,07 %	-0,78 %	-0,33 %	-0,10 %	0,80 %	-0,17 %	-0,50 %	0,20 %	-0,55 %	-0,33 %	0,24 %
5	0,28 %	0,90 %	-0,51 %	-0,89 %	-0,11 %	0,60 %	-0,34 %	-0,35 %	-0,29 %	-0,33 %	-0,07 %	0,03 %
6	-1,76 %	-0,43 %	-0,27 %	-0,74 %	-0,78 %	-0,49 %	-0,35 %	-0,50 %	-0,66 %	-0,21 %	0,39 %	-0,35 %
7	-3,82 %	-0,21 %	0,20 %	-0,02 %	-0,31 %	-0,32 %	0,43 %	-0,27 %	-0,01 %	-0,10 %	-0,15 %	0,13 %
8	-6,11 %	0,03 %	0,30 %	0,83 %	-0,31 %	-0,81 %	0,27 %	0,55 %	0,35 %	-0,45 %	0,22 %	-0,22 %
9	-8,94 %	1,05 %	0,29 %	1,28 %	0,33 %	-0,55 %	-0,42 %	1,13 %	-0,18 %	-0,89 %	-0,91 %	0,21 %
10	-16,28 %	-1,13 %	1,86 %	1,38 %	2,28 %	-1,22 %	0,98 %	0,52 %	0,34 %	0,61 %	0,76 %	-0,34 %

Source: Composed by the author

As can be seen from the Table 4, the distribution of average total returns between the deciles in periods between T1 and T11 has been narrowed significantly compared to the distribution of the base month, T0. During the period T1, the distribution of average total returns between the deciles varies from +1.07% to -1.13%. In next period T2, only four average total returns are on the positive side of the zero level. The average total returns vary from +1.86% to -0,98%. The difference of average total returns between the best performing decile and the worst performing decile is only

2.84 percentage points. As a comparison, the difference of average total returns between the best performing decile and the worst performing decile in base period T0 was 36.49 percentage points.

Periods T4 and T9 have the worst performances compared to other periods´ performances. All other deciles in period T4, except three deciles, have negative average total return. The average total returns vary between +2.28% and -0.78%. The rest three deciles have a slightly positive average total returns. All average total returns of period T9 are close to zero level. Three of them are on the positive side, and seven deciles have negative average total returns. The distribution of the average total returns of the deciles is only 2.57 percentage points.

All the average total returns of period T8 are extremely close to the zero. The difference between the best performing decile and the worst performing decile is only 1.01 percentage points. However, the period 11 has the smallest difference between the best performing decile and the worst performing decile which is only 0.78 percentage points. As a conclusion, the differences of the performance of the deciles are extremely small during each period. There is not necessary a huge spread between the best performing decile and the worst performing decile. Period 0 is an exception. There the difference between the deciles was remarkable.

The mean regression is the most obvious between periods T0 and T1. The extreme values of deciles 1 and 10 have moved towards the mediocre. The deciles with extreme values have not continued gaining as extreme values as they did on the base period. Altogether, the five best performing deciles in periods T0 have declined in average total return during period T1. At the same time, the five worst performing deciles have performed better during the period T1 than they did on the period T0. After the period T1, the average total returns have been similar to random walk. Based on the past performance, it is impossible to predict the following average total returns. Although, it seems that all average total returns of all deciles during the whole examination period have been very close to zero level, other on the positive side and other on the negative side.

The Figure 4 illustrates the change of average monthly total returns and the movement of deciles during the examination period. Similar to the line chart of average yearly total returns in the previous chapter, the mean regression is clearly detected between the periods T0 and T1. After this the deciles follow the same path with each other with minor exceptions.

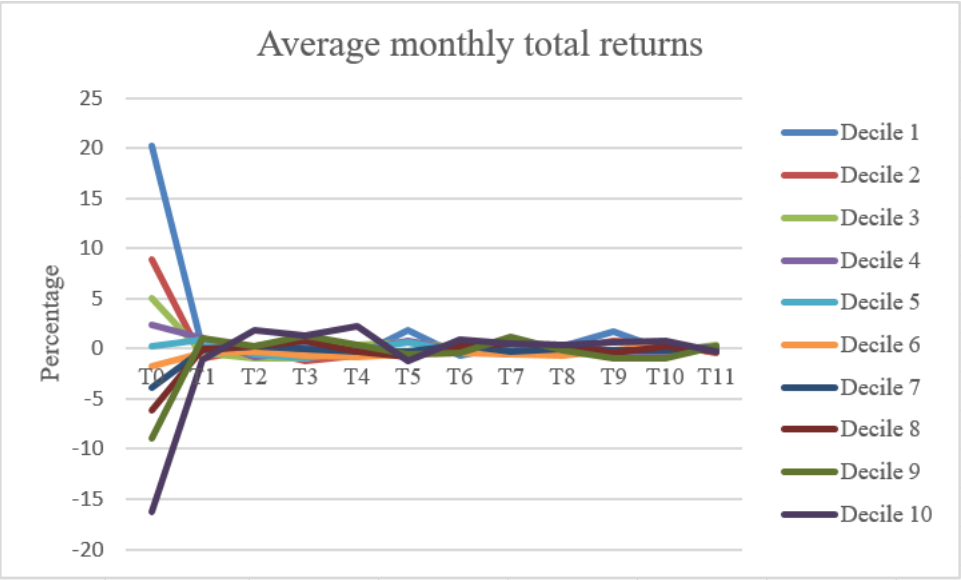

Figure 4. Average monthly total returns presented in the line chart

Source: Composed by the author

The first thing noticed from the above Figure 4 is that the decile 10 performs extremely well during the periods T0-T4. On the period T0, the increase of the average total return has been approximately 15 percentage points. After this, the average total return is the highest of all deciles in period T2. Again, during the periods T3 and T4 the average total return of the decile is better than any other deciles´ average total return. After the period T4, the decile and its average total return is mixed with other deciles. No specific information can be said about the decile´s performance based on the above figure.

The decile 1 has significant decline in the average total return from period T0 to period T1. It has not hold its position as a best performing decile. During the periods T5 and T9, the decile 1 has been the best performing decile but during the rest of the periods, its performance has been mediocre. No other significant changes can be detected from the figure. None of the deciles has hold their position, and they are all blend together with only small differences in the average total returns.

3.3 Decile movement analysis

After the thorough research of yearly and monthly average total returns, the author will look closer to the movement of each decile during the examination period. First, the author focuses on the table of the movement of the deciles of yearly average total returns. After this, the table of the movement of the deciles of monthly average total returns is taken under a closer investigation. The tables below have similar appearance than the tables on the two previous chapters have. The difference with the tables below compared to the previous tables is that the percentage points of total returns are replaced with ranking numbers. The decile with a highest total return has ranking number 1. The decile with lowest average total return gets ranking number 10. The ranking has been made based on the deciles performance on the base period, T0. The author aims to find out how has the ranking of each decile changed during the examination period, have the deciles hold their positions and are clear patterns present regarding the movement of deciles.

Table 5. Movement of the deciles of the average yearly total returns

Decile	T0	T1	T2	T3	T4	T5
1	1	4	1	2	2	2
2	2	10	5	4	4	8
3	3	9	7	7	8	6
4	4	7	4	6	3	7
5	5	6	10	10	9	10
6	6	5	9	5	6	9
7	7	3	8	8	10	5
8	8	8	6	9	5	4
9	9	2	3	3	7	3
10	10	1	2	1	1	1

Source: Composed by the author

As can be seen from the Table 5, the deciles 1 and 10 have extremely successful performance through the whole examination period. The decile 1 has hold its position as on one of the best performing deciles. After the base period, it has been ranked once as a fourth best performing decile (T1), once as the best performing decile (T2) and three times as the second best

performing decile (T3-T5). At the same time, the decile 10 has performed even better than the decile 1. It has shifted from the worst performing decile to one of the best performing deciles. The decile 10 has been ranked four times as a best performing decile, and once as a second best performing decile after the base period.

The performance of the other eight deciles has been mediocre and none of them has hold their position of the base period. None of them has had a clear direction concerning the development of the deciles. For example, the decile 7 has been the third, fifth, eighth, tenth best performing decile during the whole examination period. Another example is the decile 2. It has performed as the fourth, fifth, eighth and tenth best position. Surprisingly, the decile 5 has been ranked as worst performing decile three times during the examination period. It has not hold its position as a mediocre performing decile but instead its average total returns have been lowest of all deciles.

One reason for why the ranking between the deciles has been variable during the examination period, is that the differences of the average total returns have been relatively minor. Even a small difference in the average total return can affect to the ranking of that decile. As seen from the Table 3 of the average yearly total returns in the chapter 3.1, the differences in the average total returns can be as small as 0.01 percentage points. During the base period, the average total returns had remarkable variation between the deciles. The ranking was more straight forward during the base period. It seems impossible to forecast the future ranking of a decile based on the past performance.

Next, the author examines the movement of the deciles of the average monthly total returns. The results are presented in the Table 6.

Table 6. Movement of the deciles of the average monthly total returns

Decile	T0	T1	T2	T3	T4	T5	T6	T7	T8	T9	T10	T11
1	1	4	6	4	9	1	10	4	4	1	7	2
2	2	9	5	10	8	6	4	10	7	2	2	10
3	3	8	10	9	2	3	5	7	3	8	8	1
4	4	1	9	6	4	2	6	9	5	9	9	3
5	5	3	8	8	5	4	7	6	9	6	5	6
6	6	7	7	7	10	7	8	8	10	5	3	9
7	7	6	4	5	7	5	2	5	6	4	6	5
8	8	5	2	3	6	9	3	2	1	7	4	7
9	9	2	3	2	3	8	9	1	8	10	10	4
10	10	10	1	1	1	10	1	3	2	3	1	8

Source: Composed by the author

Unlike on the table of movement of the deciles of the average yearly total returns, the ranking of deciles in the table of average monthly total returns seems to follow a path that is similar to random walk. None of the deciles shows a clear direction of a movement. The only decile that seems to perform slightly better than other deciles is decile 10. Other deciles have shifted places with each other randomly during the whole examination period. The decile 1 has not had as good development as the decile 1 had in the Movement of the deciles of yearly average total returns - table above.

The decile 10 seems to have a better ranking during the examination period than other deciles on average. It has been ranked five times as the best performing decile, and it has couple of other top rankings as well during the periods T7-T9. There are also a few bottom rankings in periods T1, T5 and T11. Over all, it has performed well and even better than other deciles but still a good ranking in the previous period does not mean a similar ranking in the future. The performance of the decile 10 is very similar with the performance of the decile 10 in the table of movement of the deciles of average yearly total returns, above.

Other deciles do not have clear direction concerning their movement. Again, the author concludes that the main reason for this is that the differences between the performance of the deciles are extremely minor. Even a small change in the average total return can shift the decile from one ranking to another. The differences in the average total returns have been remarkable only in the base period. After this, the differences have been smaller.

3.4 Regression analysis

This part focuses on the regression analysis executed by the special regression analysis tool in Microsoft Excel. An ordinary least squares approach has been applied in the regression analysis. The regression analysis provides information of how much the performance of previous period affects to the performance of the following period. In other words, is it possible to predict the future based on the past. (For clarification, the regression in this paragraph does not refer to the concept of mean regression.) The test is done for both yearly average total returns and for monthly average total returns. The regression analysis provides a lot of information about the two chosen variables (periods) but the author focuses on the p-value and (adjusted) R square, R^2. The p-value helps to determine the significance of the results. The author presents charts with equations which include the R square value. In the text, the author also provides the adjusted R square value because the adjusted R square is a modified version of R square that has been adjusted for the number of predictors in the model. The adjusted R square is always lower than R square but the difference of these two values does not differ a lot. When making a survey on the regression analysis, the author noticed that the R square is used in the equations and the adjusted R square is used in the text part. The author follows this allocation in the following paragraphs. The confidence level in all regression analysis is 95%.

The regression analyzes for yearly average total returns are executed from the Table 3 “Average yearly total returns”. The author makes five regression analyzes comparing the two periods and the regression between them. The author presents the results of three analyzes in the following paragraphs. The three regression analyzes presented in paragraphs are regression between periods T0 and T1, regression between periods T3 and T4, and regression between periods T4 and T5. The regression analysis of T0 and T1 is presented because the author wants to demonstrate what is the result when the average total returns of two periods have a huge variation. T3 and T4 have the highest regression between them, and finally T4 and T5 are an example of mediocre regression between two periods.

The first regression analysis is made from years T0 and T1 presented in the Figure 5. The difference between the average total returns in periods T0 and T1 is the most significant than between any other periods. The distribution of average total returns in period T0 varies between +85.68% and -50.87%. In the period T1 the distribution is between +14.27% and -4.51%. In this regression analysis, the adjusted R square is 0.184 which means that 18.4% of the variation between the average total returns can be explained by the regression. (The R square is 27.5% which is shown in the equation of the figure below). The p-value is 0.12 which is more than 0.05, indicating that it is not statistically significant (since the confidence level is set as 95%). It can be inferred that the average total returns of the period T0 does not have a remarkable impact on the average total returns of the period T1.

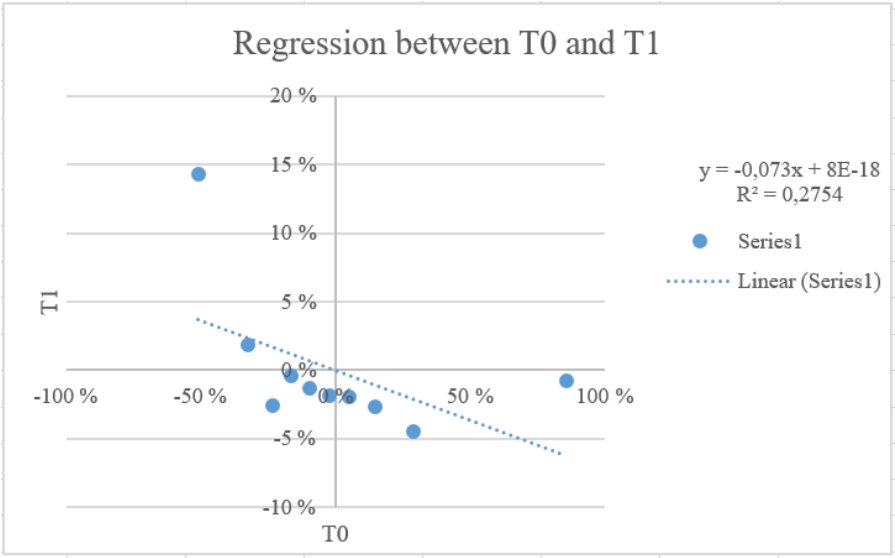

Figure 5. Regression analysis between years T0 and T1

Source: Composed by the author

The author presents the regression between periods T3 and T4 because the regression is highest between these two periods than between any other periods. As can be seen from the Figure 6 below, most of the plots are gathered around the trend line. In this regression analysis, the adjusted R square is 0.795 which means that 79.5% of the variation between the average total returns can be explained by the regression. (The R square is 81.8%). The p-value is only 0.0003

which is less than 0.05, indicating that it is statistically significant (since the confidence level is set as 95%). It can be inferred that the average total returns of the period T3 have a remarkable impact on the average total returns of the period T4. With these periods, the distribution of average total returns is very similar. In period T3 the distribution is between +6.90 and -3.06. In period T4 the distribution is between +4.02 and -2.14.

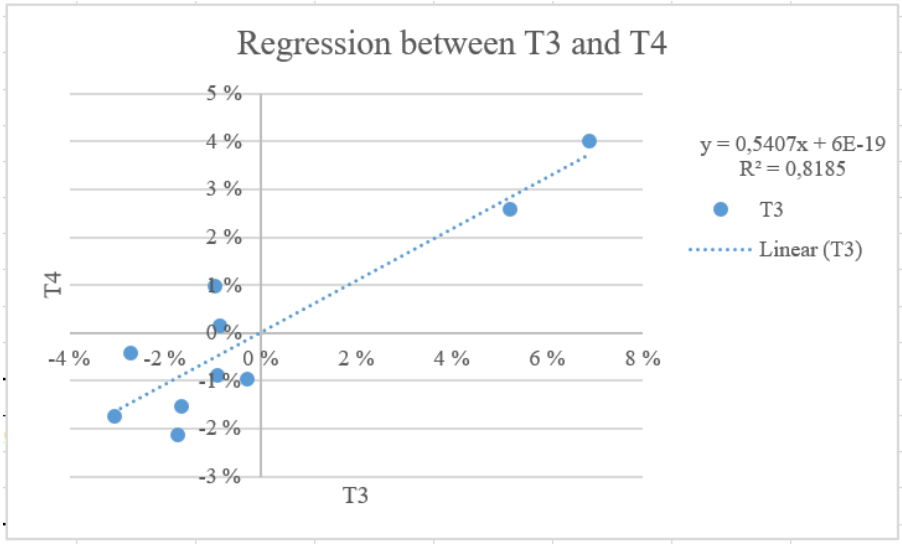

Figure 6. Regression analysis between years T0 and T1

Source: Composed by the author

The regression between T4 and T5 is mediocre in the Figure 7. 49.1% of the variation between the periods, can be explained with regression. The R square is 54.7 per cent. The other half of variation is a result of random walk. The p-value is 0.01 which is less than 0.05 with a confidence level of 95%. This indicates that the result is statistically significant. Although, the amount of information gained from the regression analysis is not remarkable. Using the results of T4 as a base for making predictions of the average total returns for period T5 is a gamble. Only the half of the results of period T5 are based on the results of T4.

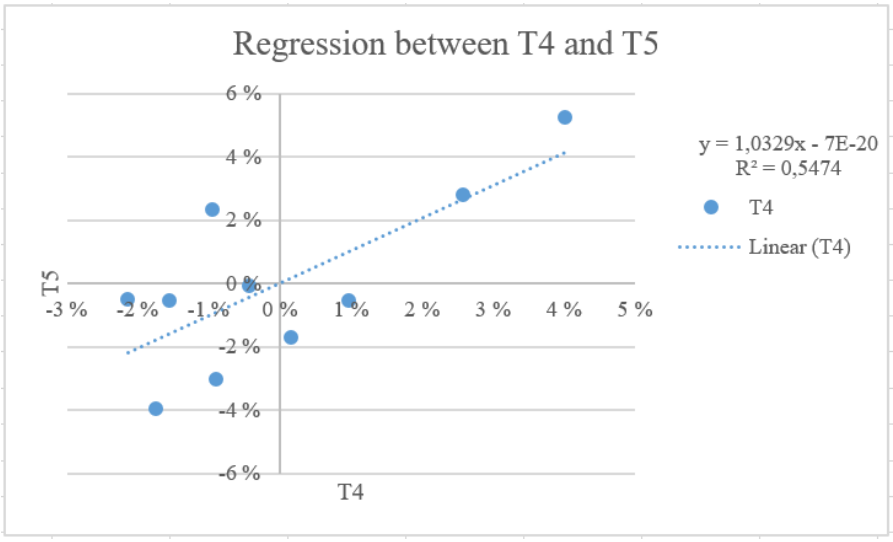

Figure 7. Regression analysis between years T4 and T5

Source: Composed by the author

The next paragraphs will illustrate the regression between the average monthly total returns. The regression analyzes are executed from the data of Table 4 Average monthly total returns. The author executes altogether eleven regression analyzes but only three regression analyzes are presented: regression between T0 and T1, regression between T2 and T3, and finally regression between T10 and T11. Most of the regression analyzes do not appear to have regression between the average total returns of two periods. Only one regression analysis shows remarkable regression, and two regression analyzes have mediocre regression. The confidence level is again 95%.

Altogether, eight regression analyzes show no regression between two periods. The author presents regression analysis between T0 and T1 which is one of the analyzes that have the lowest regression. The adjusted R square is -12%. The negative adjusted R square value is possible but it is rare. The p-value is 0.85 which indicates that the regression between T0 and T1 is almost nonexistent. As can be seen from the Figure 8 below, the plots are all over the scale. No clear direction or trend is seen on the figure.

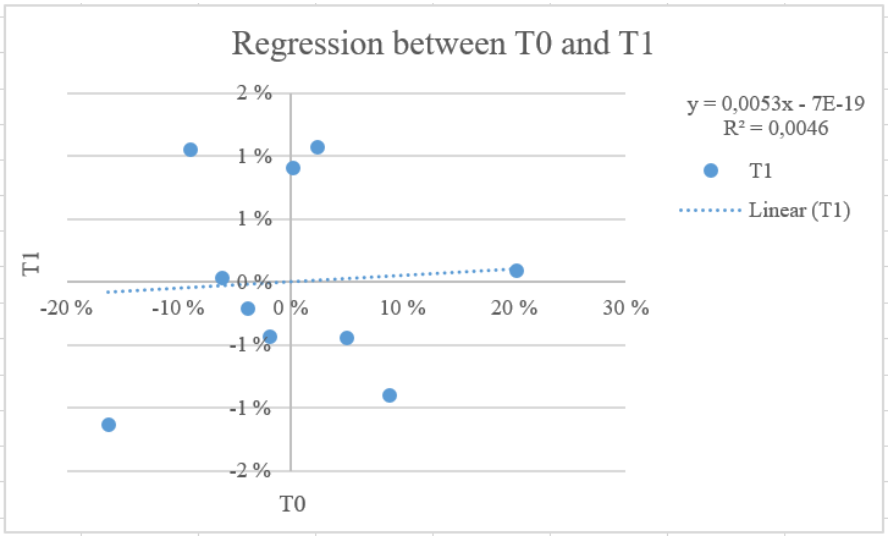

Figure 8. Regression analysis between months T0 and T1

Source: Composed by the author

Only two regression analyzes show mediocre regression between two periods. The Figure 9 illustrated the regression between months T2 and T3 which is one of those two. As can be seen from the figure, some of the plots are close to the trend line but some of them are not. This illustrates well the mediocre regression between T2 and T3. The adjusted R square is 48.5%. The p-value is 0.02 which indicates that some regression exists based on the regression analysis.

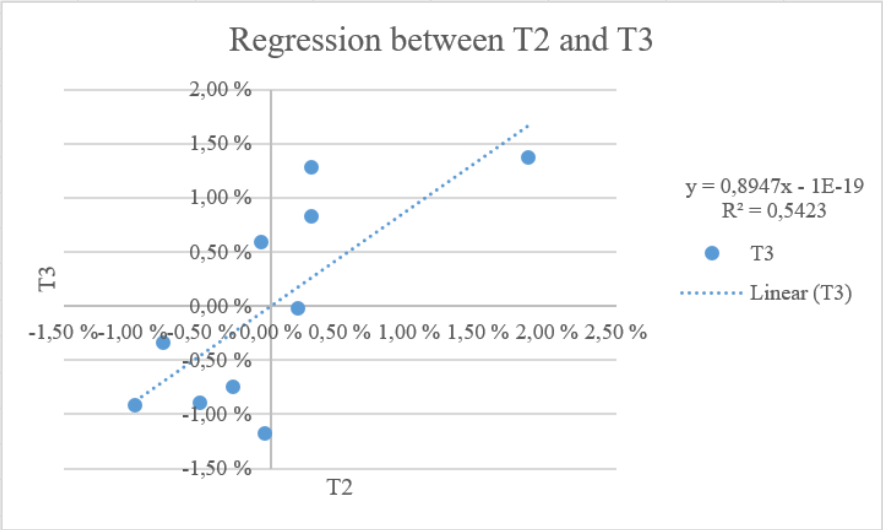

Figure 9. Regression analysis between months T2 and T3

Source: Composed by the author

The regression analysis shows remarkable regression between months T10 and T11 in the Figure 10. This is the only analysis that shows signs that the performance of previous period affects to the performance of the following period. As can be seen from the Figure 10, the dots are close to the trend line. Not much dispersion can be detected between the dots and the trend line. The adjusted R square of this regression analysis is 68.9% and R square is 72.4%. The p-value is only 0.002 which infers that the result is statistically significant when the confidence level is set as 95%.

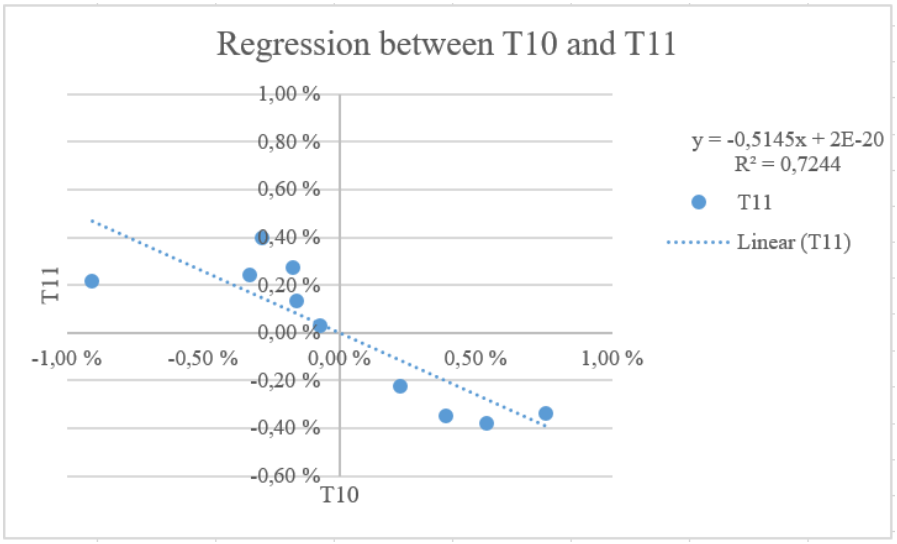

Figure 10. Regression analysis between months T10 and T11

Source: Composed by the author

3.5 Discussion

The discussion part focuses on the main findings of the four research parts. The author will explain the main reasons behind the results. The final paragraph will sum up the results with the thesis´s research objective.

The first part of the research was to examine if mean regression was detected in the yearly average total returns. As could be seen from the illustrative table of average yearly total returns and from the line chart figure, the mean regression was detected between the years T0 and T1, and especially with the deciles 1 and 10. All the deciles regressed towards the mean between T0 and T1 but with the deciles 1 and 10 the change was significant. The decile 1 did not continued performing as good as it did base on the year T0, the base year. The author expected that the average total returns of decile 1 would be more mediocre but she did not expect as sharp decline as could be seen from the illustrative table and figure. Overall, the decile 1 did show signs of mean

regression. The mean regression was not significant after the base year but the decile did manage to hold its position as one of the best performing deciles through the whole examination period.

The performance of decile 10 was also exceptional compared to the performance of other deciles. The decile showed significant mean regression between the years T0 and T1. The increase of average total return from T0 to T1 was significant. After the year T1, the decile continued to perform as one of the best deciles. Again, the author expected to detect mean regression with the decile 10 but the sharp increase in the beginning and the excellent performance after that were more than expected.

Special emphasis was put on the dynamics of the deciles 1 and 10 as those were the deciles that had the biggest changes of all deciles. The other deciles between those extremes did not show as significan. The author did not do any risk adjustments with the data although she was aware that the higher the return, the bigger the risk included. The stocks and their total returns in deciles 1 and 10 were the riskiest ones. Based on the research results, if an investor had constructed a portfolio with the stocks of decile 10, the investor would have gained exceptional returns. Again, based on the research results of decile 1, if an investor had taken a short position on the portfolio constructed with the stocks of decile 1, the investor would have gained exceptional returns with this investment strategy. But only during the year 0.

In the chapter of regression to the mean phenomenon, the author presented a research conducted by Pzena. Regression to the mean by the S&P 500 index quintiles -figure by Pzena. Pzena demonstrated regression to the mean phenomenon by making a sample of the S&P 500 index for 20 years' period. The data was divided into five quintiles based on each company´s return on equity. When the author was comparing the results, it was expected that the line chart figures of average total returns would have similar shape than the regression to the mean by the S&P 500 index quintiles -figure presented in the literature review. It was surprising for the author that the deciles did not hold their positions trough the examination period like they did on the Pzena´s research. Instead, the lines, presenting the deciles, were all mixed in together in the figure composed by the author. The author concludes that if the data had been collected in a different way, the results and illustrative figures would have been different than the current results, and would have had similar shape than the figure by Pzena. It might be also possible that the financial crisis has shaped the dynamics of the data.

The author´s thoughts about the monthly average total returns were similar with the discussion above about the yearly average total returns. The deciles 1 and 10 showed significant signs of mean regression between the months T0 and T1. After the mean regression between those months, the deciles 1 and 10 did not show clear signs of mean regression. Their performance was mixed with the performance of other deciles and the line chart figure of average monthly total returns did not show any significant performance of any deciles through the rest of the examination period.

The decile movement analysis strengthened the results of the average yearly and monthly total returns analysis. The decile movement analysis of average yearly total returns showed that the decile 1 and decile 10 had higher rankings than other deciles. In many situations, the best position and the second position belonged to these two deciles. The ranking of other deciles between the 1 and 10 did not show any interesting results in terms of the regression to the mean phenomenon.

The decile movement analysis of monthly average total returns confirmed the good performance of decile 10. It had several top positions during the whole examination period. Even tough, the differences with the average total returns were not significant, the decile movement analysis showed the differences in terms of the ranking of the deciles. The author, expected better performing from decile 1 but as stated earlier the differences between the total returns of the deciles were so minor that even a small change had an impact on the ranking.

The final part of the research was the regression analysis. This analysis was executed to show if the performance of one period had an impact on the performance of the following period. In the most cases of the both analysis, yearly and monthly, the regression did not appear. In the yearly regression analysis two cases showed regression, two cases did not have regression and in one case the regression was mediocre. In the monthly analysis, eight cases showed no regression, one case had regression and in two cases the regression was mediocre. The conclusion was that the future performance cannot be predicted based on the past performance. This supports the efficient market hypothesis.

Summing up the research results with the thesis´s objective, it appears that regression to the mean phenomenon was slightly stronger in terms of yearly perspective than in terms of monthly perspective. The mean regression was detected in the extreme values of the whole examination periods. This was something the author was expecting since according to the definition of

regression to the mean phenomenon the mean regression usually appears with extreme values. In author´s opinion, the mean regression can be used to gain abnormal returns but in minor range. The abnormal returns would have gained if an investor had taken a short position on the first decile of base period, and if an investor had invested in the last decile of the base period. Other deciles did not show mean regression, and based on that the gain of abnormal returns would have not been possible with those deciles.

3.6 Suggestions

This study represents relatively clear and straightforward method to analyze regression to the mean phenomenon. However, for the possible future researches of regression to the mean phenomenon, the author has a couple of suggestions. First, as stated earlier in the chapter of limitations of the research, the author´s original idea was to collect data of the companies present in the S&P 500 index from each particular year. With the Thomson ONE, the author was unable to get this kind of data. For the future researches, the author recommends to use different program if the purpose is to collect stock data from each particular year. One suggested program to use for this purpose would be Thomson Reuters Eikon.

Secondly, the author used Microsoft Excel for modifying and analyzing the data. The platform was easy and quick to use since the author had earlier experience about the platform. The original data downloaded to Excel was enormous but still the program was able to run the data. For the future, the author recommends to analyze the data with Rstudio or Stata. These two programs were suggested for the author to use in this thesis research. Because the author was not familiar with these programs and was not able to get access to them, she decided not to use them. The author was aware that the programs would have been more suitable to research the current topic. The author believes that these programs would offer more detailed information about the regression to the mean phenomenon.

The time frame used in this thesis research was 15 years, from 2000 to 2015. The author considered that the time frame was large enough to get valid and reliable data but for more thorough

research the author suggests to extend the time frame to 25 years or more. The wider time frame includes more economic change which would offer more distribution to the data. One also must take into account that the wider is the time frame, the higher is the turnover of the companies entering and leaving the index. This relies heavily on the availability of data and the computing power of the software used for analysis.

CONCLUSIONS

The aim of this study was to examine the regression to the mean phenomenon and its appearance in the stock market as well as to examine whether it was possible to gain abnormal returns with the phenomenon. The aim was first approached thorough a literature review and after that by research with the S&P 500 index as a data base. Methodologically, a statistical analysis was chosen as the best tool for examine regression to the mean phenomenon. For the data, the author used companies and their stocks in the Standard and Poor´s 500 index from the Standard and Poor´s index family. The author concluded that 15 year-period was long enough to detect possible regression to the mean from monthly and yearly data. By having two different periods, the author got detailed and specific information about the possible mean regression. The analysis was split into four parts: mean regression in the average yearly total returns of the stocks, mean regression in the average monthly total returns of the stocks, decile dynamics analysis and regression analysis.

Below are the most important findings of the research:

1) Deciles 1 and 10 showed clear mean regression between periods T0 and T1. The performance of other deciles was horizontal and no mean regression was detected as a result of average yearly total returns test.
2) After the average monthly total returns test, the mean regression was the most obvious between periods T0 and T1 with deciles 1 and 10. During the rest of the examination period, the deciles with extreme values have not continued gaining as extreme values as they did on the base period.
3) Concerning the decile movement analysis with average yearly total returns, the deciles 1 and 10 were ranked as the best performing deciles. The performance of the other eight deciles was mediocre and none of them hold their position of the base period nor did they had a clear direction concerning the development of the deciles.

4) The ranking of deciles in the decile movement analysis of average monthly total returns seems to follow path that is similar to random walk. None of the deciles shows a clear direction of a movement.
5) The results of regression analyzes of average yearly total returns did not have a clearly detected path. Two analyzes showed that the performance of previous period affects to the performance of the following period. Two analyzes did not show correlation between performances of two periods, and one analysis showed mediocre correlation.
6) Altogether, eight regression analyzes of average monthly total returns did not show correlation between the performance of two periods. It could be stated that was not possible to predict the performance of the following period based on the performance of previous period.

The hypothesis was that there would be mean regression in the stocks of the S&P 50 index. The study results showed that there is mean regression detected in the stocks of the S&P 500 index. Although, the mean regression did not concern the whole research data. Especially the worst performing stocks seemed to perform better when comparing two periods. Based on the study results it could be a profitable investment idea to gather a stock portfolio of the stocks that have performed worse, and take a short position on them. The study results showed that it is quite likely that those will perform better on the following period and the possibility for gaining profit is likely. The author did not study the impact of transaction costs and taxes to the possible gain.

The main problem with the thesis at hand was related to data collection. The original idea was to collect stock data in the S&P 500 index in each specific year so that each year, between 2000 and 2015, the data would have contained all the companies in the index in that specific year. During the data collection process, the author was unable to find this kind of information from the sources available to her. The current portfolio of companies was composed based on the company listing in the S&P 500 index on 29.03.2017. When making the list of companies in a specific year, the companies that did not exist in the index on that time were deleted from the portfolio. Due to this procedure, the amount of companies over a certain year varies between 410 and 505. Due to this the portfolio of companies in each year and the average total returns are not comparable with the index return of that year. One of the original ideas was to compare the return of the portfolio deciles with the index return.

For future research of regression to the mean phenomenon and its existence in the stock market, the author suggests to collect stock data from each particular year to get more precise information about the companies in the index if the purpose is to use the index as a case study. Secondly, the author states to use more advanced programs to calculate and analyze the regression to the mean phenomenon since those would offer more precise and detailed information about current topic as those deliver much higher degree of automation and therefore help to mitigate the efforts of manual operations which were relatively time consuming in Excel. Thirdly, the time frame could be extended to 25 years or more. The wider time frame would include more economic change which would offer more distribution to the data. One also must take into account that the wider is the time frame, the higher is the turnover of the companies entering and leaving the index. This relies heavily on the availability of data and the computing power of the software used for analysis.

REFERENCES

Barnet, A., van der Pols, J. and Dobson, A. (2005). Regression to the mean: what it is and how to deal with it. - *International Journal of Epidemiology*, Vol 34, pp. 215-220

Biggest Stock Exchanges In The World. (2016). Worldatlas.com http://www.worldatlas.com/articles/biggest-stock-exchanges-in-the-world.html (9.3.2017)

Bryman, A. and Bell, E. (2015). Business Research Methods, 4th revised ed. New York: Oxford University Press.

Company Overview of Standard & Poor's Financial Services LLC. Bloomberg L.P. http://www.bloomberg.com/research/stocks/private/snapshot.asp?privcapId=698396 (2.3.2017)

Geier, B. (2015). What Did We Learn From the Dotcom Stock Bubble of 2000? http://time.com/3741681/2000-dotcom-stock-bust/ (3.3.2017)

Google Finance. S&P 500 (INDEXSP:. INX). https://www.google.com/finance?q=INDEXSP:.INX (2.3.2017)

Granger, C. and Timmermann, A. (2004). Efficient market hypothesis and forecasting. International *Journal of Forecasting*. Vol 20, pp. 15-27

Graziano, A. and Raulin, M. (2004). Research methods: A process of inquiry. 5th revised ed. Boston: Allan & Bacon.

Havemann, J. (2009). The Financial Crisis of 2008: Year In Review 2008 https://global.britannica.com/topic/Financial-Crisis-of-2008-The-1484264 (3.3.2017)

Jensen, M. C. (1978). Some Anomalous Evidence Regarding Market Efficiency. - Journal of Financial Economics, Vol 6, Nos 2/3, pp. 95-101

Kahneman, D. (2011). Thinking, fast and slow. *s.l.* Farrar, Straus and Giroux.

KL Univerisity. (nd). Research Methodology. https://www.kluniversity.in/arp%5Cuploads%5C2095.pdf (3.4.2017)

Kothari, C. (2004) Research methodology. Methods & Techniques. 2nd revised ed. New Delhi: New Age International (P) Ltc.

Malkiel, B. (2003). The efficient market hypothesis and its critics http://www.vixek.com/Efficient%20Market%20Hypothesis%20and%20its%20Critics%20-%20Malkiel.pdf (13.3.2017)

Miller, M., Muthuswamy, J. and Whaley, R. (1994). Mean reversion of Standard & Poor´s 500 index basis changes: Arbitrage-included or statistical illusion?. -The Journal of Finance, Vol 49, No. 2

Murstein, B. (2003). Regression to the mean: One of the Most Neglected but Important Concepts in the Stock Market. -*The journal of Behavioral Finance*. Vol. 4, No. 4, pp. 234-237.

Organizing Your Social Sciences Research Paper: Quantitative Methods. (2017). University of Southern California. http://libguides.usc.edu/writingguide/quantitative (3.4.2017)

Primary, secondary, and tertiary sources. (2017). Virginia Polytechnic Institute and State University. http://www.lib.vt.edu/help/research/primary-secondary-tertiary.html (4.4.2017)

Pzena, R. (1995). Behavioral Biases and Investment Research. http://www.cfapubs.org/doi/pdf/10.2469/cp.v1995.n7.7 (8.3.2017)

Samsa, G. (2014). A new critique of the traditional method for empirically estimating the returns associated with strategies of stock investment: impact of fixing the investment horizon http://www.bentley.edu/sites/www.bentley.edu.centers/files/samsa.pdf (13.3.2017)

Sauer, D. and Chen, C. (1996). Mean revision in the United Kingdom stock market and its implications for a profitable trading strategy. -Journal of Business Finance & Accounting, Vol 23, issue 9-10, pp 1379-1395.

Siegel, J. (2008). Stock market. http://www.econlib.org/library/Enc/StockMarket.html (11.3.2017)

S&P U.S. Indices Methodology. S&P Dow Jones Indices, 2017a. https://us.spindices.com/documents/methodologies/methodology-sp-us-indices.pdf (2.3.2017)

S&P 500. S&P Dow Jones Indices. (2017b). http://us.spindices.com/indices/equity/sp-500 (29.03.2017)

The role of a stock exchange. (2014). London Stock Exchange plc. http://www.londonstockexchange.com/news/learningcentre/theroleofastockexchange/theroleofastockexchange.htm (10.3.2017)

Thomson Reuters. Thomson ONE
http://financial.thomsonreuters.com/en/products/tools-applications/trading-investment-tools/thomson-one-investment-research-tools.html (6.4.2017)

Urrutia, J. (1995). Tests of random walk and market efficiency for Latin American emerging equity markets. -*The Journal of Financial Research*. Vol 18. No. 3. p. 299-309.

APPENDICES

Appendix 1. List of stocks and companies in the S&P 500 index on the 29.03.2017

Company name	Ticker symbol	Company name	Ticker symbol
3M Co	MMM-N	Iron Mountain Inc	IRM-N
ABBOTT LABORATORIES	ABT-N	J B Hunt Transport Services Inc	JBHT-O
AbbVie Inc	ABBV-N	J M Smucker Co	SJM-N
Accenture PLC	ACN-N	Jacobs Engineering Group Inc	JEC-N
Activision Blizzard Inc	ATVI-O	JOHNSON & JOHNSON	JNJ-N
Acuity Brands Inc	AYI-N	Johnson Controls International PLC	JCI-N
Adobe Systems Inc	ADBE-O	JPMorgan Chase & Co	JPM-N
Advance Auto Parts Inc	AAP-N	Juniper Networks Inc	JNPR-N
Advanced Micro Devices Inc	AMD-O	KANSAS CITY SOUTHERN	KSU-N
AES Corp	AES-N	Kellogg Co	K-N
Aetna Inc	AET-N	KEYCORP	KEY-N
Affiliated Managers Group Inc	AMG-N	Kimberly-Clark Corp	KMB-N
Aflac Inc	AFL-N	Kimco Realty Corp	KIM-N
Agilent Technologies Inc	A-N	Kinder Morgan Inc	KMI-N
Air Products and Chemicals Inc	APD-N	KLA-Tencor Corp	KLAC-O
Akamai Technologies Inc	AKAM-O	Kohls Corp	KSS-N
Alaska Air Group Inc	ALK-N	Kraft Heinz Co	KHC-US
Albemarle Corp	ALB-N	Kroger Co	KR-N
Alexandria Real Estate Equities Inc	ARE-N	L Brands Inc	LB-N
Alexion Pharmaceuticals Inc	ALXN-O	L3 Technologies Inc	LLL-N
Allegion PLC	ALLE-N	LABORATORY CORPORATION OF AMERICA HOLDINGS	LH-N
Allergan plc	AGN-N	Lam Research Corp	LRCX-O
Alliance Data Systems Corp	ADS-N	Leggett & Platt Inc	LEG-N
Alliant Energy Corp	LNT-N	Lennar Corp	LEN-N
Allstate Corp	ALL-N	Leucadia National Corp	LUK-N
Alphabet Inc	GOOG-O	Level 3 Communications Inc	LVLT-N
Alphabet Inc	GOOGL-O	Lincoln National Corp	LNC-N
Altria Group Inc	MO-N	LKQ Corp	LKQ-O
Amazon.com Inc	AMZN-O	Lockheed Martin Corp	LMT-N
Ameren Corp	AEE-N	Loews Corp	L-N
American Airlines Group Inc	AAL-O	Lowe's Companies Inc	LOW-N
American Electric Power Company Inc	AEP-N	LyondellBasell Industries NV	LYB-N
American Express Co	AXP-N	M&T Bank Corp	MTB-N
American International Group Inc	AIG-N	Macerich Co	MAC-N

Appendix 1 continued

American Tower Corp	AMT-N
American Water Works Company Inc	AWK-N
Ameriprise Financial Inc	AMP-N
AmerisourceBergen Corp	ABC-N
Ametek Inc	AME-N
Amgen Inc	AMGN-O
Amphenol Corp	APH-N
Anadarko Petroleum Corp	APC-N
Analog Devices Inc	ADI-O
Anthem Inc	ANTM-N
AON PLC	AON-N
Apache Corp	APA-N
Apartment Investment and Management Co	AIV-N
Apple Inc	AAPL-O
Applied Materials Inc	AMAT-O
Archer Daniels Midland Co	ADM-N
Arconic Inc	ARNC-N
Arthur J Gallagher & Co	AJG-N
Assurant Inc	AIZ-N
AT&T Inc	T-N
Autodesk Inc	ADSK-O
Automatic Data Processing Inc	ADP-O
AutoNation Inc	AN-N
Autozone Inc	AZO-N
AvalonBay Communities Inc	AVB-N
Avery Dennison Corp	AVY-N
Baker Hughes Inc	BHI-N
Ball Corp	BLL-N
Bank of America Corp	BAC-N
Bank of New York Mellon Corp	BK-N
Baxter International Inc	BAX-N
BB&T Corp	BBT-N
Becton Dickinson and Co	BDX-N
Bed Bath & Beyond Inc	BBBY-O
Berkshire Hathaway Inc	BRK'B-N
Best Buy Co Inc	BBY-N
Biogen Inc	BIIB-O
BlackRock Inc	BLK-N
Boeing Co	BA-N
BorgWarner Inc	BWA-N
Macy's Inc	M-N
Mallinckrodt plc	MNK-N
Marathon Oil Corp	MRO-N
Marathon Petroleum Corp	MPC-N
Marriott International Inc	MAR-O
Marsh & McLennan Companies Inc	MMC-N
Martin Marietta Materials Inc	MLM-N
Masco Corp	MAS-N
MasterCard Inc	MA-N
Mattel Inc	MAT-O
McCormick & Company Inc	MKC-N
McDonald's Corp	MCD-N
McKesson Corp	MCK-N
Mead Johnson Nutrition Co	MJN-N
Medtronic PLC	MDT-N
Merck & Co Inc	MRK-N
Metlife Inc	MET-N
Mettler-Toledo International Inc	MTD-N
Michael Kors Holdings Ltd	KORS-N
Microchip Technology Inc	MCHP-O
Micron Technology Inc	MU-O
Microsoft Corp	MSFT-O
Mid-America Apartment Communities Inc	MAA-N
Mohawk Industries Inc	MHK-N
Molson Coors Brewing Co	TAP-N
Mondelez International Inc	MDLZ-O
Monsanto Co	MON-N
Monster Beverage Corp	MNST-O
Moody's Corp	MCO-N
MORGAN STANLEY	MS-N
Mosaic Co	MOS-N
Motorola Solutions Inc	MSI-N
Murphy Oil Corp	MUR-N
Mylan NV	MYL-O
Nasdaq Inc	NDAQ-O
National Oilwell Varco Inc	NOV-N
Navient Corp	NAVI-O
NetApp Inc	NTAP-O
Netflix Inc	NFLX-O
Newell Brands Inc	NWL-N

Appendix 1 continued

Boston Properties Inc	BXP-N
Boston Scientific Corp	BSX-N
Bristol-Myers Squibb Co	BMY-N
Broadcom Ltd	AVGO-O
Brown-Forman Corp	BF'B-N
C R Bard Inc	BCR-N
C.H. Robinson Worldwide Inc	CHRW-O
CA Inc	CA-O
Cabot Oil & Gas Corp	COG-N
Campbell Soup Co	CPB-N
Capital One Financial Corp	COF-N
Cardinal Health Inc	CAH-N
Carmax Inc	KMX-N
Carnival Corp	CCL-N
Caterpillar Inc	CAT-N
CBOE Holdings Inc	CBOE-O
CBRE Group Inc	CBG-N
CBS Corp	CBS-N
Celgene Corp	CELG-O
Centene Corp	CNC-N
CENTERPOINT ENERGY INC	CNP-N
CenturyLink Inc	CTL-N
Cerner Corp	CERN-O
CF Industries Holdings Inc	CF-N
Charles Schwab Corp	SCHW-N
Charter Communications Inc	CHTR-O
Chesapeake Energy Corp	CHK-N
Chevron Corp	CVX-N
Chipotle Mexican Grill Inc	CMG-N
Chubb Ltd	CB-N
Church & Dwight Co Inc	CHD-N
Cigna Corp	CI-N
Cimarex Energy Co	XEC-N
Cincinnati Financial Corp	CINF-O
Cintas Corp	CTAS-O
Cisco Systems Inc	CSCO-O
Citigroup Inc	C-N
Citizens Financial Group Inc	CFG-N
Citrix Systems Inc	CTXS-O
Clorox Co	CLX-N
CME Group Inc	CME-O
CMS Energy Corp	CMS-N
Newfield Exploration Co	NFX-N
Newmont Mining Corp	NEM-N
News Corp	NWSA-O
News Corp	NWS-O
NextEra Energy Inc	NEE-N
Nielsen Holdings PLC	NLSN-N
Nike Inc	NKE-N
NiSource Inc	NI-N
Noble Energy Inc	NBL-N
Nordstrom Inc	JWN-N
Norfolk Southern Corp	NSC-N
Northern Trust Corp	NTRS-O
Northrop Grumman Corp	NOC-N
NRG Energy Inc	NRG-N
Nucor Corp	NUE-N
NVIDIA Corp	NVDA-O
O'Reilly Automotive Inc	ORLY-O
Occidental Petroleum Corp	OXY-N
Omnicom Group Inc	OMC-N
ONEOK Inc	OKE-N
Oracle Corp	ORCL-N
PACCAR INC	PCAR-O
Parker-Hannifin Corp	PH-N
Patterson Companies Inc	PDCO-O
Paychex Inc	PAYX-O
PayPal Holdings Inc	PYPL-O
Pentair PLC	PNR-N
People's United Financial Inc	PBCT-O
PepsiCo Inc	PEP-N
PerkinElmer Inc	PKI-N
Perrigo Company PLC	PRGO-N
Pfizer Inc	PFE-N
PG&E Corp	PCG-N
Philip Morris International Inc	PM-N
PHILLIPS 66	PSX-N
Pinnacle West Capital Corp	PNW-N
Pioneer Natural Resources Co	PXD-N
PNC Financial Services Group Inc	PNC-N
PPG Industries Inc	PPG-N
PPL Corp	PPL-N
Praxair Inc	PX-N
Priceline Group Inc	PCLN-O

Appendix 1 continued

Coach Inc	COH-N	Principal Financial Group Inc	PFG-N
Coca-Cola Co	KO-N	Procter & Gamble Co	PG-N
Cognizant Technology Solutions Corp	CTSH-O	Progressive Corp	PGR-N
Colgate-Palmolive Co	CL-N	Prologis Inc	PLD-N
Comcast Corp	CMCSA-O	Prudential Financial Inc	PRU-N
Comerica Inc	CMA-N	Public Service Enterprise Group Inc	PEG-N
Conagra Brands Inc	CAG-N	PUBLIC STORAGE	PSA-N
Concho Resources Inc	CXO-N	PulteGroup Inc	PHM-N
CONOCOPHILLIPS	COP-N	PVH Corp	PVH-N
Consolidated Edison Inc	ED-N	Qorvo Inc	QRVO-O
Constellation Brands Inc	STZ-N	Qualcomm Inc	QCOM-O
Cooper Companies Inc	COO-N	Quanta Services Inc	PWR-N
Corning Inc	GLW-N	Quest Diagnostics Inc	DGX-N
Costco Wholesale Corp	COST-O	Ralph Lauren Corp	RL-N
Coty Inc	COTY-N	Range Resources Corp	RRC-N
Crown Castle International Corp	CCI-N	Raymond James Financial Inc	RJF-N
CSRA Inc	CSRA-N	Raytheon Co	RTN-N
CSX Corp	CSX-O	Realty Income Corp	O-N
Cummins Inc	CMI-N	Red Hat Inc	RHT-N
CVS Health Corp	CVS-N	Regency Centers Corp	REG-N
D.R. Horton Inc	DHI-N	Regeneron Pharmaceuticals Inc	REGN-O
Danaher Corp	DHR-N	Regions Financial Corp	RF-N
Darden Restaurants Inc	DRI-N	Republic Services Inc	RSG-N
Davita Inc	DVA-N	Reynolds American Inc	RAI-N
Deere & Co	DE-N	Robert Half International Inc	RHI-N
DELPHI AUTOMOTIVE PLC	DLPH-N	Rockwell Automation Inc	ROK-N
Delta Air Lines Inc	DAL-N	Rockwell Collins Inc	COL-N
DENTSPLY SIRONA Inc	XRAY-O	Roper Technologies Inc	ROP-N
Devon Energy Corp	DVN-N	Ross Stores Inc	ROST-O
Digital Realty Trust Inc	DLR-N	Royal Caribbean Cruises Ltd	RCL-N
DISCOVER FINANCIAL SERVICES	DFS-N	Ryder System Inc	R-N
Discovery Communications Inc	DISCK-O	S&P Global Inc	SPGI-N
Discovery Communications Inc	DISCA-O	Salesforce.com Inc	CRM-N
DISH Network Corp	DISH-O	SCANA Corp	SCG-N
Dollar General Corp	DG-N	Schlumberger NV	SLB-N
Dollar Tree Inc	DLTR-O	Scripps Networks Interactive Inc	SNI-O
Dominion Resources Inc	D-N	Seagate Technology PLC	STX-O
Dover Corp	DOV-N	Sealed Air Corp	SEE-N
Dow Chemical Co	DOW-N	SEMPRA ENERGY	SRE-N
Dr Pepper Snapple Group Inc	DPS-N	Sherwin-Williams Co	SHW-N
DTE Energy Co	DTE-N	Signet Jewelers Ltd	SIG-N

Appendix 1 continued

Duke Energy Corp	DUK-N
Dun & Bradstreet Corp	DNB-N
E I du Pont de Nemours and Co	DD-N
E*TRADE Financial Corp	ETFC-O
Eastman Chemical Co	EMN-N
Eaton Corporation PLC	ETN-N
eBay Inc	EBAY-O
Ecolab Inc	ECL-N
EDISON INTERNATIONAL	EIX-N
Edwards Lifesciences Corp	EW-N
Electronic Arts Inc	EA-O
Eli Lilly and Co	LLY-N
Emerson Electric Co	EMR-N
Entergy Corp	ETR-N
Envision Healthcare Corp	EVHC-N
EOG Resources Inc	EOG-N
EQT Corp	EQT-N
Equifax Inc	EFX-N
Equinix Inc	EQIX-O
EQUITY RESIDENTIAL	EQR-N
Essex Property Trust Inc	ESS-N
Estee Lauder Companies Inc	EL-N
Eversource Energy	ES-N
Exelon Corp	EXC-N
Expedia Inc	EXPE-O
Expeditors International of Washington Inc	EXPD-O
Express Scripts Holding Co	ESRX-O
Extra Space Storage Inc	EXR-N
Exxon Mobil Corp	XOM-N
F5 Networks Inc	FFIV-O
Facebook Inc	FB-O
Fastenal Co	FAST-O
FEDERAL REALTY INVESTMENT TRUST	FRT-N
FedEx Corp	FDX-N
Fidelity National Information Services Inc	FIS-N
FIFTH THIRD BANCORP	FITB-O
Firstenergy Corp	FE-N
Fiserv Inc	FISV-O
FLIR Systems Inc	FLIR-O
Simon Property Group Inc	SPG-N
Skyworks Solutions Inc	SWKS-O
SL Green Realty Corp	SLG-N
Snap-On Inc	SNA-N
Southern Co	SO-N
Southwest Airlines Co	LUV-N
Southwestern Energy Co	SWN-N
Stanley Black & Decker Inc	SWK-N
Staples Inc	SPLS-O
Starbucks Corp	SBUX-O
State Street Corp	STT-N
Stericycle Inc	SRCL-O
Stryker Corp	SYK-N
SunTrust Banks Inc	STI-N
Symantec Corp	SYMC-O
Synchrony Financial	SYF-N
Synopsys Inc	SNPS-O
Sysco Corp	SYY-N
T. Rowe Price Group Inc	TROW-O
Target Corp	TGT-N
TE Connectivity Ltd	TEL-N
TechnipFMC PLC	FTI-N
Tegna Inc	TGNA-N
Teradata Corp	TDC-N
Tesoro Corp	TSO-N
Texas Instruments Inc	TXN-O
Textron Inc	TXT-N
Thermo Fisher Scientific Inc	TMO-N
Tiffany & Co	TIF-N
Time Warner Inc	TWX-N
TJX Companies Inc	TJX-N
Torchmark Corp	TMK-N
Total System Services Inc	TSS-N
Tractor Supply Co	TSCO-O
TransDigm Group Inc	TDG-N
Transocean Ltd	RIG-N
Travelers Companies Inc	TRV-N
TripAdvisor Inc	TRIP-O
Twenty-First Century Fox Inc	FOX-O

Appendix 1 continued

Flowserve Corp	FLS-N
Fluor Corp	FLR-N
FMC Corp	FMC-N
Foot Locker Inc	FL-N
Ford Motor Co	F-N
Fortive Corp	FTV-N
Fortune Brands Home & Security Inc	FBHS-N
Franklin Resources Inc	BEN-N
Freeport-McMoRan Inc	FCX-N
Gap Inc	GPS-N
Garmin Ltd	GRMN-O
General Dynamics Corp	GD-N
General Electric Co	GE-N
General Mills Inc	GIS-N
General Motors Co	GM-N
Genuine Parts Co	GPC-N
GGP Inc	GGP-N
Gilead Sciences Inc	GILD-O
Global Payments Inc	GPN-N
Goldman Sachs Group Inc	GS-N
Goodyear Tire & Rubber Co	GT-O
H & R Block Inc	HRB-N
Halliburton Co	HAL-N
Hanesbrands Inc	HBI-N
Harley-Davidson Inc	HOG-N
Harris Corp	HRS-N
Hartford Financial Services Group Inc	HIG-N
Hasbro Inc	HAS-O
HCA Holdings Inc	HCA-N
HCP Inc	HCP-N
Helmerich and Payne Inc	HP-N
Henry Schein Inc	HSIC-O
Hershey Co	HSY-N
Hess Corp	HES-N
Hewlett Packard Enterprise Co	HPE-N
Hologic Inc	HOLX-O
Home Depot Inc	HD-N
Honeywell International Inc	HON-N
Hormel Foods Corp	HRL-N
Host Hotels & Resorts Inc	HST-N
HP Inc	HPQ-N
Humana Inc	HUM-N
Twenty-First Century Fox Inc	FOXA-O
Tyson Foods Inc	TSN-N
U.S. BANCORP	USB-N
UDR Inc	UDR-N
Ulta Beauty Inc	ULTA-O
Under Armour Inc	UA-N
Under Armour Inc	UAA-N
Union Pacific Corp	UNP-N
United Continental Holdings Inc	UAL-N
United Parcel Service Inc	UPS-N
United Rentals Inc	URI-N
United Technologies Corp	UTX-N
UnitedHealth Group Inc	UNH-N
Universal Health Services Inc	UHS-N
UNUM GROUP	UNM-N
Valero Energy Corp	VLO-N
Varian Medical Systems Inc	VAR-N
Ventas Inc	VTR-N
Verisign Inc	VRSN-O
Verisk Analytics Inc	VRSK-O
Verizon Communications Inc	VZ-N
Vertex Pharmaceuticals Inc	VRTX-O
VF Corp	VFC-N
Viacom Inc	VIAB-O
Visa Inc	V-N
VORNADO REALTY TRUST	VNO-N
Vulcan Materials Co	VMC-N
W W Grainger Inc	GWW-N
Wal-Mart Stores Inc	WMT-N
Walgreens Boots Alliance Inc	WBA-O
Walt Disney Co	DIS-N
Waste Management Inc	WM-N
Waters Corp	WAT-N
WEC Energy Group Inc	WEC-N
Wells Fargo & Co	WFC-N
Welltower Inc	HCN-N
Western Digital Corp	WDC-O
Western Union Co	WU-N
WestRock Co	WRK-N
Weyerhaeuser Co	WY-N
Whirlpool Corp	WHR-N
Whole Foods Market Inc	WFM-O

Appendix 1 continued

Huntington Bancshares Inc	HBAN-O	Williams Companies Inc	WMB-N
IDEXX Laboratories Inc	IDXX-O	Willis Towers Watson PLC	WLTW-O
Illinois Tool Works Inc	ITW-N	Wyndham Worldwide Corp	WYN-N
Illumina Inc	ILMN-O	Wynn Resorts Ltd	WYNN-O
Incyte Corp	INCY-O	Xcel Energy Inc	XEL-N
Ingersoll-Rand PLC	IR-N	Xerox Corp	XRX-N
Intel Corp	INTC-O	Xilinx Inc	XLNX-O
Intercontinental Exchange Inc	ICE-N	XL Group Ltd	XL-N
International Business Machines Corp	IBM-N	Xylem Inc	XYL-N
International Flavors & Fragrances Inc	IFF-N	Yahoo! Inc	YHOO-O
International Paper Co	IP-N	Yum! Brands Inc	YUM-N
Interpublic Group of Companies Inc	IPG-N	Zimmer Biomet Holdings Inc	ZBH-N
Intuit Inc	INTU-O	Zions Bancorp	ZION-O
Intuitive Surgical Inc	ISRG-O	Zoetis Inc	ZTS-N
Invesco Ltd	IVZ-N		