School of Economics and Management

The relationship between the returns of the S&P 500 index and the Volatility Index VIX

Master thesis Finance

Name: Mike van Wees
Snr.: 2031259
Anr.: 287368
Subject: Master thesis Finance
Supervisor: Dr. Rik Frehen

Preface

This thesis is a major and important component of my master Finance at Tilburg University. In the past period I completed this research with a lot of dedication. This results in my thesis: 'The relationship between the returns of the S&P 500 and the Volatility Index VIX'.

I would like to thank my supervisor Dr. Rik Frehen, who guided me through the process, and helped me when I had problems during my thesis. During our meetings, the valuable tips from you helped me a lot and you helped me to take this thesis to a higher level with good results. Furthermore, I would like to thank my family and friends with the support throughout the thesis.

I hope you find the thesis an interesting research.

Mike van Wees,

Tilburg, September 28, 2020

Abstract

The research question during this thesis is: Is there a relationship between the returns of the S&P 500 index and the VIX, and is it possible to set up a profitable trading strategy for the S&P 500 index using the VIX? During the study, different models were compared to determine whether the VIX might have predictive power. Subsequently, a rule of thumb was drawn up based on one of the models. Then, the relationship between the value of the VIX and the returns of the S&P 500 index is analysed. Finally, an investment strategy based on the VIX is developed. It can be concluded that there is a relationship between the value of the VIX and the returns of the S&P 500 index. When the VIX is high, the returns from the S&P 500 index are extremely widespread and these returns decrease. In addition, the investment strategy tested in the thesis is very interesting for investors as this strategy achieves significantly higher returns, compared to when the investor always stays in the S&P 500 index. Using this strategy investors can profit from the relationship between the VIX and returns of the S&P 500 index.

Table of contents

1. Introduction

During this thesis, it is investigated whether there is a relationship between the returns of the S&P 500 index and the value of the volatility index VIX. The S&P 500 index is a stock market index that measures the performance of the largest 500 companies, which are listed on the stock market in the United states. This index is one of the most important equity indices in the world (Wang, 2008).

The volatility index VIX is measured since 1993 by the Chicago Board Options Exchange (CBOE). The VIX is originally designed to measure the market expectations of 30 day volatility implied by at the money S&P 100 index options. However in 2003 the VIX is updated as a result of that the VIX measures market expectations of the near-term volatility implied by stock index option prices. The VIX is constructed to be a general estimator of the market's estimate of the S&P 500 volatility over 30 days (Blair, Poon, & Taylor, 2001).

In addition to investigating whether there is a relationship between the VIX and the returns of the S&P 500 index, the study also investigated whether the VIX has predictive power and whether the VIX can be used during the development of an investment strategy. This leads to the main question of the thesis, which is used as a guideline for the research. This main question is as follows:

Is there a relationship between the returns of the S&P 500 index and the VIX, and is it possible to set up a profitable trading strategy for the S&P 500 index using the VIX?

According to Whaley (2000) the volatility index VIX is also called the investor fear gauge. The VIX measures the implied index volatility of the S&P 500 over 30 days. Since the implied volatility is the forward looking volatility, the VIX can be seen as a predictor for uncertain times (Shaikh & Padhi, 2015). This confirms Whaley's (2000) opinion, because when the volatility and the VIX increases it is possible that fear among investors increases and that is why the VIX is also called the investor fear gauge. This opinion indicates that investors see the VIX as a fear meter, this opinion makes the research question in this thesis relevant to determine whether there is a relationship between the VIX and the returns of the S&P 500 index. By answering this question, it can be determined whether the VIX can really be seen as a fear predictor.

In addition to that is this research question economically relevant on the basis of the market on which the VIX is based. The VIX is based on the S&P 500 index, which is the largest American equity market. When it can be established on the basis of this thesis that there is a relationship between the VIX and the returns of the S&P 500 index, and this thesis also shows a strategy that ensures that returns can be increased, this is interesting for many investors, as many of them invest in the S&P 500.

Various datasets are used during this thesis. These datasets are mainly collected through WRDS, data that was not yet available through WRDS is collected through Yahoo Finance. The dataset used in the thesis is the value of the VIX between January 1993 and May 2020. In addition, the returns of the S&P 500 index are used in this thesis. The dataset regarding the returns of the S&P 500 index is also from the period January 1993 to May 2020. These datasets serve as the basis for the empirical model and to answer the research question.

The first step in the empirical model is to draw up the simple regression model. This model predicts the returns of the S&P 500 index based on the historical returns of the S&P 500. The historical returns are lagged by one month and in addition also by one day. Subsequently, the multivariate regression model is drawn up, using this model, returns are predicted based on historical returns and the VIX. Historical returns as well as the VIX are lagged by one month and also by one day. One of the four models is used for establishing the rule of thumb. Before this rule of thumb is drawn up, it is first decided which model will be used, based on the adjusted R-squared and information criteria. Ultimately, a rule on thumb is drawn up based on the returns, which are predicted using one of the four models.

Finally, an investment strategy is described in the empirical model. This strategy is based on the VIX. The investment strategy that is tested in the thesis is one based on the data from January 1993 to May 2020 of the VIX. The mean of the VIX over this period of 6,902 observations is 19.38. The standard deviation of the VIX over this period of 6,902 observations is 8.39. The investment strategy that I want to test in the thesis is to exit the market when the VIX is higher than the value of the mean plus one standard deviation. This means that I will exit the market if the VIX has a value higher than 27.77 (19.38 + 8.39) and that I enter the market again when the VIX is lower than 27.77. Subsequently, the returns between the restricted and unrestricted strategies are compared to analyse whether the strategy performs better than if the investor always remains in the market.

The simple regression and multivariate regression model were first created in the chapter results to determine which of the models performs best. It can be concluded that all values of the adjusted R-squared of the models are very low, since the independent variables of all models explains less than a percent of the variation in the dependent variable. This also makes sense since the exchange market is unpredictable. When the exchange market is predictable, it is possible for any individual to predict the market and generate big profits.

Based on the adjusted R-squared and the information criteria, I concluded that the simple regression model performs better than the multivariate regression model. This applies to the models that are lagged with one month the models that are lagged with one day. The multivariate regression model included the lagged VIX. Based on the choice of the model and the data I used, it can be concluded that the VIX does not contribute to predicting returns of the S&P 500 index.

Using the returns predicted by means of the simple regression model, I determine the following rules regarding investing in the S&P 500 index or exiting the S&P 500 index.

- I will exit the S&P 500 index when the model predicts that the returns will be -12.12 percent or lower. This means I will exit the market when the predicted return is below the mean minus one the standard deviation.
- I will enter the S&P 500 index when the models predicts that return will be 18.26 or higher. This means that I will enter the market when the predicted return is higher than the mean plus one standard deviation.

When the previously described rules are followed by an investor, this investor exits the market 42 times over the past 27 years and enters the market 45 times or buys additional assets at that time in over the past 27 years.

Finally, the investment strategy based on the VIX is tested. I test this second strategy since the first strategy was not based on the VIX and I want to set up a profitable investment strategy using the VIX. This strategy is tested on the basis of two different periods. The entire period that the VIX exists, from January 1993 to May 2020 and the period from January 2017 to May 2020. For the period 1993 to 2020 the average daily return using the investment rule is 0.07 percent, this daily return is relatively much higher than the 0.035 percent if the investor always remains in the market. For the period 1993 to 2020 the average daily return using the investment rule is 0.057 percent, this daily return is relatively much higher than the 0.044 percent if the investor always remains in the market. Based on the two outcomes in the different samples, it can be concluded, that higher returns can be achieved by means of the investment strategy based on the VIX. In comparison with the unrestricted strategy when the investor always remains in the market. This conclusion is reinforced as the results are also significant. For the period from January 1993 to May 2020, the significance level is 1 percent and the period January 2017 to May 2020, the significance level is 5 percent.

The rest of the thesis is organized as follows. Section 2 literature review, section 3 dataset en descriptive statistics, section 4 empirical model, section 5 results and section 6 conclusion.

2. Literature Review

In this chapter, the study begins by explaining the main variable of the thesis in section 1. Section 2 discusses why the main variable should be used. Moving on with section 3, stock market crashes from 1993. The last part of this chapter analyses what is already known about predicting stock market yields using historical literature.

2.1 What is the VIX

The objective of this research is to investigate the relationship between the VIX and stock market returns. The main variable of the thesis is the VIX. The volatility index VIX is measured since 1993 by the Chicago Board Options Exchange (CBOE). The VIX is originally designed to measure the market expectations of 30 day volatility implied by at the money S&P 100 index options. However in 2003 the VIX is updated as a result of that the VIX measures market expectations of the near-term volatility implied by stock index option prices. The VIX depends on the prices of a portfolio 30 calendar day S&P 500 calls and puts with weights being inversely proportional to the squared acquisition strike price (Fernandes, Medeiros, & Scharth, 2014). The VIX is an implied volatility index derived from put and call options of the S&P 500 index with maturities of 30 days (22 trading days) (Becker, Clements, & McClelland, 2009). The VIX is constructed to be a general estimator of the market's estimate of the S&P 500 volatility over 30 days (Blair, Poon, & Taylor, 2001).

2.2 Choice of VIX

The aim of this research to determine whether there is a relationship between the VIX and the returns of the S&P 500 index. It is then analysed whether it is possible to draw up a profitable investment strategy using the VIX. The main component which is used in this research to set up this strategy is the volatility index VIX. The choice of the VIX is based on the value as predictor for investors and in addition the market on which the VIX is based, namely the S&P 500 index options.

2.2.1 Investor fear gauge

According to Whaley (2000) the volatility index VIX is also called the investor fear gauge. The VIX measures the implied index volatility of the S&P 500 over 30 days. Since the implied volatility is the forward looking volatility, the VIX can be seen as a predictor for uncertain times (Shaikh & Padhi, 2015). This confirms Whaley's (2000) opinion, because when the volatility and the VIX increases it is possible that fear among investors increases and that is why the VIX is also called the investor fear gauge. This statement is supported by Sarwar (2012), his study discovered a strong relation between changes in the VIX and daily stock market returns in U.S. Brazil between 1993 and 2007. This study also concluded that the VIX responded much more aggressive to negative changes in stock market returns than that the VIX responds to positive changes in stock market returns. This suggests that the VIX is a gauge for of investors fear (Sarwar, 2012). This makes the VIX a good predictor of potential market crashes as it measures investor fears and confidence.

2.2.2 Private information among investors

While trading stocks and setting up investment strategies, having information is very important. It is easier to predict future outcomes using specific information. Previous research concludes that investors who have access to private information executes many profitable transactions as a result of the information they have at their disposal (Bushee & Goodman, 2007). As discussed earlier, the VIX is derived from call and put options of the S&P 500. As a result of that, it is possible for investors to cash in their private information as much as possible through options.

Tsai, Chiu & Wang (2015) also conducted research on the value of private information related to investing. In the end, they presented evidence regarding the VIX index options. They conclude that traders wish to act on the information that they possess in the VIX options market. When investors have specific information they are likely to choose to sell their shares using limit orders, as opposed to marketable orders (Tsai, Chiu, & Wang, 2015). Investors uses limit orders instead of market orders to limit execution price uncertainty. In addition to this conclusion, other researchers have investigated whether informed options investors predict stock returns. Chang, Hsieh & Lai have evidence from the Taiwan stock exchange market. They investigated the influence of options on predictability and concluded that the group of informed traders provided the largest predictability in the middle horizon options and the near the money options (Chang, Hsieh, & Lai, 2009). This makes it easier for this group of investors to predict future returns and makes it more likely that this group will make better investment decisions.

Earlier in this chapter it was stated that the VIX is forward looking. As a result of that, it is possible to see the VIX as a possible predictor for future returns. This makes this variable an interesting variable for investors. It is also possible that investors have private information and as a result of that perform various actions. An example is when some of the investors have private information indicating that potentially bad economic times are approaching. Investors can take advantage of this through options and gain potentially high returns. Since the VIX is an implied volatility index derived from put and call options of the S&P 500, it is possible that the VIX might change significantly due to private information among investors. If so, the VIX is a good estimator for future returns. This makes this research whether the VIX and the returns of the S&P 500 index have a relationship economically interesting for many investors and companies. In addition, it is also economically interesting to investigate whether the VIX can be used to set up an investment strategy.

2.2.3 S&P 500 Index

Besides that the VIX is a good predictor according to investors, the market on which the VIX is based, is also very interesting. The VIX is based on the S&P 500 market index. The S&P 500 index is a stock market index that measures the performance of the largest 500 companies, which are listed on the stock market in the United states. This index is one of the most important equity indices in the world and many investors and companies consider this index as one of the best representations of the stock markets in the United States (Wang, 2008).

The importance of the S&P 500 makes it economically interesting to use the VIX as the main variable in this study, as many investors are using the S&P 500 as a key index in the US equity markets. During this thesis, the predictive value of the VIX is analysed. What the VIX is and why the VIX is chosen was described in the previous section on the basis of papers.

2.3 Exchange market crashes

During this thesis it is investigated whether there is a relationship between the VIX and exchange market crashes. Since the VIX exists since 1993, there is only data available from this date. As a result of that, research in this thesis will be conducted from 1993. As of 1993, there have been three major stock U.S. market crashes. These were the dot-com bubble in 2000 and the Financial crisis in 2008 (Barro & Ursuá, 2017). The third market crash was the coronavirus stock market crash in 2020.

2.3.1 Dot-com Bubble

In the late 1990s, the internet became more accessible to more people around the world. During that period, investors focused on these internet based companies. As a result of that, the value of companies in the internet sector rose sharply. In addition, many new companies were started up and went public during this period. In 1999, 446 internet based start-ups went public and they gain an average return of 70 percent on the first day. The degree of investment in internet based companies and the desire to grow internet start-ups quickly lead to the highest ranking of the NASDAQ ever. On March 10, 2000 the NASDAQ peaked at 5048.62. This was the highest value of the NASDAQ during the Dot-com Bubble (Goonight & Green, 2010).

On March 13, the NASDAQ opens 4.5% lower. Analysts view this lower opening as a market correction, however the NASDAQ's decline continued. During April 2020 the internet index lost 19 percent of its value. The market value of internet companies declined form €1 trillion in March 2000 to only €572 billion in December 2000. The NASDAQ also collapsed in 2000 as a result of the Dot-com Bubble. At the end of 2000, NASDAQ recorded 2470.52, which is a decrease of 52 percent from March 2000 (Goonight & Green, 2010). When the bubble burst in 2000, it caused a global recession that was unexpectedly protracted in some Western countries.

2.3.2 Financial crisis

The Financial crisis can be divided into two phases. The first phase runs from August 2007 to August 2008, this limited phase stemmed from losses in one relatively small segment in the U.S. financial system, namely subprime mortgages. Despite this disruption to financial markets, caused by the subprime mortgages, the real GDP in the United States continued to rise into the second quarter of 2008 and analysts only predicted a small recession (Mishkin, 2011).

The second phase is the global financial crisis. The bankruptcy of Lehman Brothers on Monday, 15th of September is considered as the start of the global financial crisis. Since the bankruptcy of Lehman Brothers, much uncertainty has arisen among investors. This financial crisis led to a worldwide recession (Mishkin, 2011). The crisis peaked in October 2008 and ended in 2011. From 2010, the main concerns shifted from the housing market crisis to the worrisome financial positions of governments, for example Greece (D.Gibson, G.Hall, & S.Tavlas, 2012).

2.3.3. Coronavirus stock market crash

The most recent crisis is the stock market crash of 2020. With regard to this crisis, the Coronavirus appears to be the black swan. The Coronavirus has led to major losses among investors around the whole world. Large indexes lost about 10 percent of their value on March 9, 2020. The largest stock market declines since September 9, 2001 have occurred on this day. The losses on March 9, 2020 even exceed the losses in 2008 during the financial crisis due to the bankruptcy of Lehman Brothers (Daube, 2020 (working paper)). In addition to the Corona crisis, there is another aspect that contributed to stock market crash. This aspect was the oil war between Saudi Arabia and Russia. These two factors lead to the stock market crash in March 2020 and subsequent high unemployment in the United States.

Since the data that is being investigated is only available from 1993, the analysis in this thesis is done from this period. The three events have been named and described to indicate the impact on the stock market when a crash occurs. In addition, these events have also been described in order to gain insight into when data relating to the VIX must strongly change from the average value.

2.4 Predicting market returns or exchange market downturns

During this part of the literature review, it is analysed which important information with regard to predicting returns in the past has already been collected. Firstly, the relationship between volatility and stock market returns is analysed. Thereafter it is examined what is already known in the literature about the VIX as a predictor of stock market returns. Finally, the economic value of predicting stock market returns and volatility is discussed.

2.4.1 Volatility and stock market returns

In the past, a lot of research has been done into the relationship between volatility and stock market returns. French & Schwert (1987) analysed a possible relationship between stock returns and the stock market volatility. They concluded that they found evidence for a positive relationship between the market risk premium and the predictability of the volatility of stock returns. In addition, French & Schwert (1987) also found evidence for a negative relationship between unexpected stock market returns and unexpected changes in the volatility of stock market returns. They considered this as indirect evidence for a positive relationship between expected risk premiums of stock and the volatility of these stocks.

Baillie & DeGennaro (1990) stated that most asset pricing models also suggest a positive relation between portfolio's expected returns and risk. This risk is often measured using the variance. This makes sense since the investor requires reward for the risk incurred. However, Baillie & DeGennaro (1990) concluded after estimating a variety of models from daily and monthly return that the relationship between mean returns and standard deviation is weak. As a result of their results Baillie & DeGennaro (1990) suggests that investors consider other risk measures instead of the variance of standard deviation during their investment decisions.

Following on from the above paragraph, Guo & Savickas conducted further research and have additional information from previous research into volatility and stock market returns. Guo & Savickas (2012) find that the idiosyncratic stock volatility and the aggregate stock market volatility together exhibit a strong predictive power for stock return markets. Guo & Savickas (2012) also concluded that a high level of idiosyncratic volatility is usually associated with low expected future stock returns. The combination between risk and return is positive as mentioned earlier, however idiosyncratic volatility is negative related to stock market returns (Guo & Savickas, 2012).

Ang et al. (2009) have also researched the relationship between idiosyncratic volatility and future returns. The results of this research reinforce the results of the conclusion mentioned earlier. Ang et al (2009) conducted research based on 23 developed countries and concluded that stocks with recent past high idiosyncratic volatility have low future average returns around these 23 countries. In addition, this conclusion is significant in the G7 countries (Ang et al. , 2009). This conclusion is logical, since in finance idiosyncratic risk is seen as a diversifiable risk. As a result of that, investors are not compensated for idiosyncratic risk.

The literature mentioned earlier refers to the relationship between stock market returns and volatility. Important research has been conducted in the past into this relationship between volatility and stock market returns. During my thesis I will compare the relationship between the VIX and stock market returns, this is also a known relationship but based on the VIX. This thesis extends research between volatility and returns as the VIX is also volatility but of the entire S&P 500 index.

2.4.2 The VIX and stock market returns

As mentioned earlier in the literature review, the VIX is also used as a predictor for stock market returns. The VIX is known as a investors fear index. This VIX is the risk-neutral expected stock market variance for the U.S. S&P 500 options, the VIX is calculated using implied volatility from the options of the S&P 500 index (Whaley, 2000). Previous research also supports to use implied volatility and VIX as predictor instead of for example historical volatility.

Szakmary et al. (2003) conducted research into the predictive power of implied volatility. During the research, they involved many different futures markets in the research. Thirty five futures were analysed in the study, including the S&P 500 index, which is the VIX. Szakmary et al (2003) concluded that implied volatility is a better predictor of realized volatility compared to historical volatility. Since in the majority of futures markets, the prediction of the implied volatility of the realized volatility was better than the historical volatility.

Szakmary et. Al (2003) concluded that implied volatility was a good predictor of realized volatility based on thirty five different futures markets. One of these thirty five futures markets was the S&P 500. Ederington & Guan (2002) have also researched the implied volatility as a predictor. During their analyses they decided to use only data from the S&P 500 index options. Based on this data, Ederington & Guan wanted to find out whether the implied volatility is an efficient and effective predictor of future volatility.

Ederington & Guan (2002) concluded that the implied volatility has a strong predictive power. They also concluded that prediction results and efficiency results are sensitive to the forecasted horizon and when the data covers a stock market crash, for example the stock market crash of 1987 (Ederington & Guan, 2002). This conclusion confirms that the implied volatility of the S&P 500 index options is a good predictor of realized volatility.

Bekaert & Hoerova (2014) re-examined and further expand the predictive power of the volatility index VIX. Unlike to other researchers, Bekaert & Hoerova first decomposed the squared VIX into two components: the conditional variance of the stock market and the equity premium variance. The equity variance premium is the difference between the squared VIX and the conditional variance of the stock market (Variance premium = squared VIX – conditional variance of the stock market). Eventually Bekaert & Hoerova concluded that the variance premium is a significant predictor of future returns. In addition, they conclude that the conditional variance of the stock market predicts negative economic activity and that the conditional variance of stock markets has a high predictive power of financial instability (Bekaert & Hoerova, 2014).

Despite the fact that many papers mention the predictive power of the VIX and implied volatility, there are also papers with doubt about the VIX as a predictor for returns and realized volatility. Kownastki (2016) questioned how good the VIX is as a predictor for market risk. Kownastki (2016) concluded that during the most critical time periods, for example the Financial crisis in 2008, the VIX does not perform as promised. He concluded that the VIX understates realized volatility by about 180 basis points on average. He also concluded that poor timing increases possible forecast errors in the future (Kownatzki, 2016).

Based on this literature review, I still think that the VIX will be the best main variable during my research. Since the VIX is generally regarded as one of the best predictors by investors. in addition, several papers and researchers have confirmed this assumption in the literature review above. The literature described above also provides a good basis for my thesis. Several papers show that the VIX has a predictive power for stock market returns, I will test this hypothesis in the rest of the thesis.

2.4.3 Economic value of predicting volatility and stock market returns

Forecasting market returns and volatility is economically interesting for investors, using this forecasts, investors can make profits and avoid losses. Investors can use different strategies while investing.

Marqueing & Verbeek (2004) analysed the economic value of predicting stock returns and volatility. They examine several investment strategies, using data from 1970 until 2001. They conclude that for a mean-variance investor, predicting volatility is profitable, even if transaction costs are very high and short sales are not allowed during the investment strategies.

During the research in my thesis, it is tested whether there is a relationship between the VIX and returns of the S&P 500 index. Furthermore it is tested whether the VIX has predictive power of predcting of returns of the S&P 500 index. This is economically interesting for many investors since the VIX is based on the S&P 500, which is one of the most important stock exchange markets in the world. In addition, many investors see the VIX as a fear gauge and as a result of that they use the VIX during their investment decisions.

3. Dataset and descriptive statistics

Chapter 3 is divided into three different sections. The first paragraph explains the sample selection. The second paragraph describes the institutional features of the market in addition, the data sources are listed, including the datasets used in this thesis. Finally, the descriptive statistics are presented in the last paragraph.

3.1 sample selection

Before this study describes the relationship between the VIX and the returns of the S&P 500 index, a brief summary of the institutional setting is helpful to put things into perspective. In order to compare the performance of stock markets with the VIX, the data regarding the S&P 500 is being used in this thesis. The choice to make the comparison of the VIX as a predictor using data from the S&P 500 is economically interesting, since the S&P 500 consists of the largest publicly traded companies within the United States.

The data which is used in this thesis is extracted for the time period 1993 until May 2020. During this period three major financial events occurred. These events had major influence on the VIX and stock markets around the world. These three events were: The Dot-com Bubble in the late 1990s, the Financial Crisis in 2007 and 2008 and the Coronavirus stock market crash in 2020.

The data will be investigated from 1993, because the VIX is only available from this period (Whaley, 2000). In addition, it is decided to investigate the data up to and including May 2020, so that the performance of the VIX during the Coronavirus crisis in March 2020 can be included in the investigation. The starting sample consists of 6,902 observations in the period 1993 until May 2020.

3.2 Institutional features and datasets

The institutional details of the S&P 500 are provided during this section of the chapter. The chapter also identifies the data sources and datasets, which are used during the analysis.

3.2.1 Institutional features S&P 500

The S&P 500 index is a market-capitalization weighted index consisting of the 500 largest publicly traded companies in the United States. The S&P 500 index is also called Standard & poor's 500 index. The index is considered the most important United States equity index worldwide, because this index represents the largest publicly traded corporations in the United States (Wang, 2008).

The S&P 500 consists of the 500 largest publicly traded companies in the United States. Only the top 10 largest companies are listed on the Standard & Poor website. Many of the top 10 companies are technology companies or financial companies. For example, Microsoft, Apple, Facebook and JP Morgan Chase & Co. Investors prefer the S&P 500 to other US indices because this index has more stocks than, for example, the Dow Jones Industrial Average (500 versus 30).

3.2.2 Data sources and datasets

I use two different data sources to collect the data for my research. I use Wharton Research Data Service (WRDS) to get most of my datasets, namely the data from 1993 to 2019. WRDS provides leading business intelligence, research platform, and data analytics to global institutions, historical analysis and insight into the latest innovations in research (WRDS, 2020). Since I also want to include the impact of the coronavirus crisis during my analysis, I decided to also collect data from January 2020 to May 2020. This data is collected through Yahoo Finance and added to the data collected through WRDS. Yahoo Finance provides free stock quotes, up-to-date news, portfolio management, and international market data (Yahoo Finance, 2020)

During my analysis I use three different datasets. The first dataset is called CBOE S&P 500 Volatility Index (VIX) and consists of data from January 1993 through May 2020. The second dataset is called S&P 500 index daily return and also consists of data from January 1993 to May 2020. Finally, the third dataset is a combination of the two previously mentioned datasets.

3.3 Descriptive statistics

The sample consists of 6,902 observations of the two different variables. The data refers to daily data of trading days on the US stock exchange market from January 1993 through May 2020. As previously reported, the datasets are obtained through WRDS and are merged into a dataset that will be used during the analysis.

Looking at the main variable in this thesis which is the VIX, the mean is 19.38. Furthermore, the minimum value of the VIX 9.14 and the maximum value of the VIX 82.69. This maximum value is a major outlier when the mean is compared to the standard deviation, which indicates that the maximum values of the VIX are more extreme than the minimum values of the VIX. These high values mainly occur during uncertain times on the stock market.

The other variable, daily return of the S&P 500 in percentages, is collected to determine the value of the VIX using investing in the S&P 500 index. From 1993 the average daily return of the S&P 500 index is 0.035%. It is true that there might be large outliers of daily returns, which are from the minimum and maximum value. By means of this data, I try to determine an investment strategy using the VIX, which outperforms a strategy when the investor always remain in the market. The performance of this strategy is compared on the basis of the returns of the restricted and unrestricted strategy.

Figure 1: Summary statistics
The summary statistics regarding the data from January 1993 until May 2020 are shown in the table below. The variables are Volatility Index VIX and the daily return of the S&P 500 index.

Variable	Obs	Mean	Std. Dev.	Min	Max
VIX	6,902	19.377	8.394	9.14	82.69
DailyReturn	6,902	0.00035	.012	-.12	.116

4. Empirical model

In this chapter of the thesis, the empirical models used during the analysis are described. In addition to the empirical models, the estimators are also described during this part of the thesis. First, the simple regression model is described and elaborated. One variable is then added to this simple regression model to test whether it helps predicting the value of this added variable. Thereafter, a rule of the thumb is added to the analysis and the rule of thumb is further explained during this part of the thesis. Finally, an investment strategy based on the VIX is tested during the analysis.

4.1 Simple regression model

During the analysis of the thesis, I predict the value of the VIX by comparing two different models. One model excluding the lagged VIX and another model including the lagged VIX. To determine the predictive power of the VIX, a reference point must first be drawn up for comparison. This reference point is an estimation of the returns on the S&P 500 index (Rm_t) based on the historical returns on the S&P 500 index (Rm_{t-1}). To estimate returns based on historical returns, historical returns are regressed on t+1. The term used for the lag t+1 is one month. In addition, a model is also regressed based on a one day lag. This leads to the following simple regression model:

$$Rm_t = \beta_0 + \beta_1 * Rm_{t-1} + u_t \quad \text{(I)}$$

4.2 Multivariate regression model

After using the simple regression model as a reference point, the multivariate regression model is created. This model consists of the simple regression model as mentioned earlier, with a new variable added, namely the VIX. The lagged VIX is added to the model to assess whether the VIX has predictive power for the S&P 500's returns. The term used for the lag t+1 of the VIX is also one month. In addition, a model is also regressed based on a one day lag. This leads to the following multivariate regression model:

$$Rm_t = \beta_0 + \beta_1 * Rm_{t-1} + \beta_2 * VIX_{t-1} + u_t \quad \text{(II)}$$

Ultimately, this model is used to analyse whether the VIX has predictive power regarding the returns of the S&P 500. To analyse whether the VIX has predictive power, the adjusted R-squared of the simple regression model and multiple regression model are compared. In addition, is it also possible to look at the loading of the coefficient of the VIX and it is also possible to use the information criteria in Stata to select the best model in Stata. When the coefficient of the VIX is relatively high, this coefficient contributes more to predicting returns in comparison when the loading on the coefficient is low.

4.3 Rule of thumb

During the analysis of the thesis, a rule of thumb is drawn up based on one of the four models. Before the rule is drawn up, a choice must first be made between the simple regression models and the multivariate regression models. The model which will be chosen is more valuable as a predictor of the returns of the S&P 500 index in comparison with the other models.

The choice of model is based on two different aspects. This concerns the adjusted R-squared per model and the information criteria. Based on these aspects it is decided which model is used during the rule of thumb.

Subsequently, the values are calculated in Excel by using one of the models. These values are then plotted for the graphical display. Finally, a rule is drawn up. Based on the values, which are known by the model, this rule indicates moments when to enter and invest in the S&P 500 index, as well as times to exit and sell the investments.

4.4 Investment strategy using the VIX

After testing the predictive power of VIX using the four different models, the VIX is used to test an investment strategy, which is based on the value of the VIX. Before the investment strategy is tested, a scatter plot of the value of the VIX and the returns of the S&P 500 index is added. This scatter plot indicates whether there might be a relationship between the VIX and the performance of the returns of the S&P 500 index. In addition, a regression is also added which predicts returns based on the lagged VIX.

The investment strategy that is tested in the thesis is one based on the data from January 1993 to May 2020 of the VIX. The mean of the VIX over this period of 6,902 observations is 19.38 (see Figure 1: Summary statistics). The standard deviation of the VIX over this period of 6,902 observations is 8.39 (see Figure 1: Summary statistics). The investment strategy which I want to test in the thesis is to get out of the market when the VIX is higher than the value of the mean plus one standard deviation. This means that I will exit the market if the VIX has a value higher than 27.77 (19.38 + 8.39) and that I will return and enter the market again when the VIX is lower than 27.77.

The next step is to compare the returns of the restricted strategy and the unrestricted strategy. During this step, the returns for the restricted strategy and the unrestricted strategy are first calculated. Subsequently, the average returns of both strategies is calculated. Based on these average returns I conclude whether the strategy based on the VIX achieve better results than if the investor always stays in the market. Finally, the results are tested whether they are significant.

5. Results

During this part of thesis, the analysis is performed and the results are analysed. First, the simple regression models are created. Then the multivariate regression models are created. Thereafter, this chapter of the thesis explains the rule of the thumb, this rule of thumb is based on a the simple regression model or on a multivariate regression model. The choice of the model is based on the predictive power and the use in practice. Finally, an investment strategy based on the VIX is performed and tested.

5.1 Simple regression model

During the thesis four different models are compared, this concerns two simple regression models and two multivariate regression models. First, the simple regression model is described and interpreted. Using the simple regression model, returns of the S&P 500 index are predicted based on lagged historical returns of the S&P 500 index. This leads to the following simple regression model:

$$Rm_t = \beta_0 + \beta_1 * Rm_{t-1} + u_t \qquad \text{(I)}$$

Rm_{t-1} are the historical returns of the S&P 500 index lagged by 1 month. By means of Stata the daily data was first adjusted to monthly data. Subsequently, these monthly data were lagged with 1 month using Stata. The lagged historical returns of the S&P 500 index are chosen as estimator to provide a reference point to compare with the multivariate regression model in which the lagged VIX is included.

Based on the lagged historical returns of the S&P 500 index, the table (see Figure 2) below is generated in Stata by predicting the returns of the S&P 500 index based on the one month lagged returns of the S&P 500 index.

Figure 2: Simple regression model based on historical returns lagged with one month

VARIABLES	(1) Rmt
Rm_Lagged	0.03603
	(0.065)
Constant	0.00659***
	(2.79)
Observations	328
Adjusted R-squared	- 0.0018

This table shows the result of one regression from the period May 1993 until January 2020. The dependent variable is return of the market (S&P 500 index) and the independent variable is the one month lag of return of the market (S&P 500 index). The numbers in the parentheses are the t-statistics. *, **, *** indicates the significance levels of 0.10, 0.05 and 0.01.

Based on the table in figure 2 it is possible to make two different interpretations:

- 1 Percentage point increase in the monthly lagged return of the S&P 500 index results in an expected increase of 3.603 percentage points in the return of the S&P 500 index.
- If the lagged returns of the S&P 500 index is zero, expected return of the S&P 500 index is 0.659 percentage point.

β_1 is not statistically significant since the t-statistic is 0.65. This means that it is not possible to conclude that this coefficient is statistically significant even at 10 percent. The constant factor is statistically significant at 1 percent, since the t-statistic is 2.79. This means that the coefficient is statistically significant at 1 percent.

The adjusted R-squared of the simple regression model in figure 2 is −0.0018 which is even negative, this indicates the variation in the model is not much explained by the model. This makes sense as the stock market is unpredictable. When the market is predictable, it is possible for any individual to predict the market and generate big profits.

As mentioned earlier, the model is regressed based on historical returns lagged with one month. I also decide to regress the regression model bases on historical returns with a lag of one day to determine which form of lag allows the model to perform best. The results of this regression are shown in the table below (see figure 3).

Figure 3: Simple regression model based on historical returns lagged with one day

VARIABLES	(1) Rmt
Rm_Lagged	-0.09756***
	(-8.14)
Constant	0.00039***
	(2.73)
Observations	6,901
Adjusted R-squared	0.0094

This table shows the result of one regression from the period May 1993 until January 2020. The dependent variable is return of the market (S&P 500 index) and the independent variable is the one day lag of return of the market (S&P 500 index). The numbers in the parentheses are the t-statistics. *, **, *** indicates the significance levels of 0.10, 0.05 and 0.01.

Based on the table in figure 3 it is possible to make two different interpretations:

- 1 Percentage point increase in the daily lagged return of the S&P 500 index results in an expected decrease of 9.756 percentage points in the return of the S&P 500 index.
- If the lagged returns of the S&P 500 index is zero, expected return of the S&P 500 index is 0.0039 percentage point.

β_1 is statistically significant since the t-statistic is -8.14. This means that it is possible to conclude that this coefficient is statistically significant even at 1 percent. The constant factor is statistically significant at 1 percent, since the t-statistic is 2.73. This means that the coefficient is statistically significant at 1 percent.

The adjusted R-squared of the simple regression model in figure 3 is 0.0094, this indicates the variation in the model is not much explained by the model. This makes sense as the stock market is unpredictable. When the market is predictable, it is possible for any individual to predict the market and generate big profits.

5.2 Multivariate regression model

In addition to the simple regression model, the multivariate regression model is also described and interpreted. Using the multivariate regression model, returns of the S&P 500 index are predicted based on lagged historical returns of the S&P 500 index and lagged values of the VIX. This leads to the following Multivariate regression model:

$$Rm_t = \beta_0 + \beta_1 * Rm_{t-1} + \beta_2 * VIX_{t-1} + u_t \qquad \text{(II)}$$

As mentioned earlier with the simple regression model, Rm_{t-1} are the historical returns of the S&P 500 index lagged by 1 month. With Stata, the daily data is first adjusted to monthly data. Subsequently, these monthly data is lagged by 1 month with Stata. In addition to the simple regression model, the lagged VIX is added as a coefficient to predict the return of the market. VIX_{t-1} is the value of the Volatility Index (VIX) lagged by 1 month. The daily data of the VIX is first adjusted to monthly data with use of Stata. Subsequently, these monthly data from the VIX were lagged with 1 month by means of Stata. This results in the multivariate regression model.

Using the lagged historical returns of the S&P 500 index and the lagged values of the Volatility Index (VIX), the model (see Figure 4) below is generated in Stata by predicting the returns of the S&P 500 index based on lagged returns of the S&P 500 index and lagged values of the VIX.

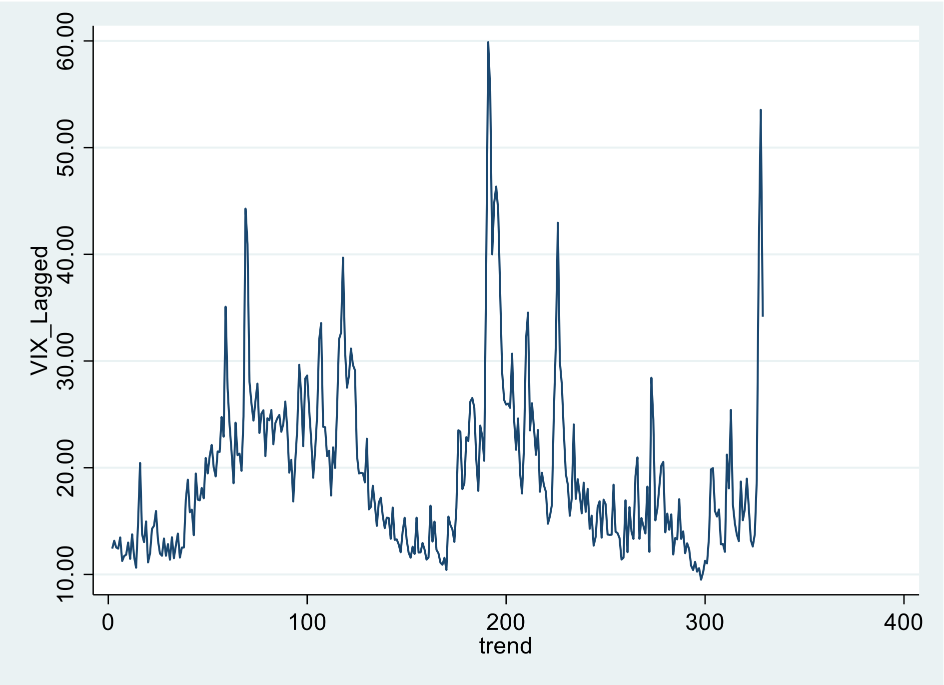
Figure 4: Multivariate regression model based on historical returns and VIX lagged with one month

VARIABLES	(1) Rmt
Rm_Lagged	0.05587
	(0.92)
VIX_Lagged	0.00026
	(0.81)
Constant	0.00134
	(0.19)
Observations	328
Adjusted R-squared	- 0.0028

This table shows the result of one regression from the period May 1993 until January 2020. The dependent variable is return of the market (S&P 500 index) and the independent variables are the one month lag of return of the market (S&P 500 index) and the one month lag of the value of the VIX. The numbers in the parentheses are the t-statistics. *, **, *** indicates the significance levels of 0.10, 0.05 and 0.01.

It might be possible that there is a unit root in the data related to the VIX. To test whether there is a unit root, the Augmented Dickey-Fuller unit root test by means of Stata is used. A time series process is non stationary if this series contains a unit root. The sequence of the dataset of the VIX is tested by means of the Augmented Dickey-Fuller unit test. The sequence is first shown visually in figure 5.

Figure 5: Graph of the sequence of the lagged VIX



This figure provides a graphical representation of the pattern of the VIX. The data is monthly data from the VIX from January 1993 to May 2020.

To test whether there is a unit root in the dataset with the lagged VIX, the Augmented Dickey-fuller unit root test in Stata is performed. The outcome of the augmented Dickey-Fuller unit root test is shown in Figure 6.

Figure 6: Augmented Dickey-Fuller unit root test

Dickey-Fuller test for unit root   Number of obs = 327

	Test Statistic	Interpolated Dickey-Fuller 1% Critical Value	5% Critical Value	10% Critical Value
Z(t)	-5.324	-3.454	-2.877	-2.570

MacKinnon approximate p-value for Z(t) = 0.0000

D.VIX_lagged	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
VIX_lagged L1.	-.1644484	.0308892	-5.32	0.000	-.2252165	-.1036803
_cons	3.259282	.6470597	5.04	0.000	1.986328	4.532236

This model represents the outcome of the test for a possible unit root. The unit root test was performed using the Augmented Dickey-Fueller test. The test was conducted on data related to the VIX during the period January 1993 to May 2020.

The results of the Augmented Dickey-Fuller unit root test are shown in Figure 5. It turns out that the value of the t-statistic is -5,324 and it is to the left of -3,454. -3.454 Is the value for 1 percent critical value. In addition, the P-value is 0.0. These two results conclude that we can reject the null hypothesis. This means that we do not have a unit root in the dataset of the lagged VIX and that the process is not non stationary. If there was a unit root I could solve this by using the first difference of the lagged VIX, however this is not necessary as there is no unit root.

Since there is no unit root in the dataset of the lagged VIX, we can use the table as shown in figure 4. Based on this table it is possible to make three different interpretations:

- 1 Percentage point increase in the lagged return of the S&P 500 index results in an expected increase of 5.587 percentage points in the return of the S&P 500 index, holding other factors fixed.
- 1 Unit increase in the lagged VIX results in an expected increase of 0.026 percentage point in the return of the S&P 500 index, holding other factors fixed.
- If the lagged returns of the S&P 500 index and the lagged VIX are equal to zero, expected return of the S&P 500 index is 0.134 percentage point.

β_1 is not statistically significant since the t-statistic is 0.92. This means that it is not possible to conclude that this coefficient is statistically significant even at 10 percent. β_2 is not statistically significant since the t-statistic is 0.81. This means that it is not possible to conclude that this coefficient is statistically significant even at 10 percent. The constant factor is also statistically not significant at 10 percent, since the t-statistic is 0.19. This means that al coefficients of the multivariate regression model are not statistically significant.

The adjusted R-squared of the multivariate regression model in figure 4 is -0.0028 which is even negative, this indicates the variation in the model is not much explained by the model. This is the same as with the simple regression model since the stock market is not predictable.

As mentioned earlier, the model is regressed based on historical returns lagged with one month. I also decide to regress the regression model bases on historical returns with a lag of one day to determine which form of lag allows the model to perform best. The results of this regression are shown in the table below (see figure 7).

Figure 7: Multivariate regression model based on historical returns and VIX lagged with one day

VARIABLES	(1) Rmt
Rm_Lagged	-0.09461***
	(-7.84)
VIX_Lagged	0.00003**
	(1.98)
Constant	-0.00026
	(-0.74)
Observations	6,901
Adjusted R-squared	0.0098

This table shows the result of one regression from the period May 1993 until January 2020. The dependent variable is return of the market (S&P 500 index) and the independent variables are the one day lag of return of the market (S&P 500 index) and the one day lag of the value of the VIX. The numbers in the parentheses are the t-statistics. *, **, *** indicates the significance levels of 0.10, 0.05 and 0.01.

Based on the table in figure 7 it is possible to make three different interpretations:

- 1 Percentage point increase in the lagged return of the S&P 500 index results in an expected decrease of 9.461 percentage points in the return of the S&P 500 index, holding other factors fixed.
- 1 Unit increase in the lagged VIX results in an expected increase of 0.003 percentage point in the return of the S&P 500 index, holding other factors fixed.
- If the lagged returns of the S&P 500 index and the lagged VIX are equal to zero, expected return of the S&P 500 index is -0.026 percentage point.

β_1 is statistically significant since the t-statistic is -7.84. This means that it is possible to conclude that this coefficient is statistically significant at 1 percent. β_2 is statistically significant at 5 percent significance level since the t-statistic is 1.98. The constant factor is statistically not significant at 10 percent, since the t-statistic is -0.74.

The adjusted R-squared of the multivariate regression model in figure 7 is 0.0098 which is very low, this indicates that the variation in the model is not much explained by the model.

5.3 Choice of the model

In this part of the thesis, it is decided which model is used to determine the rule of thumb. The choice is made between the four models mentioned earlier in this chapter. When making the choice of the model, the adjusted R-squared and the information criteria in Stata are taken into account.

5.3.1 Adjusted R-Squared of the model

The adjusted R-squared captures what percentage of variation in the dependent variable is explained by the independent variables. The adjusted R-squared of the first simple regression model (Figure 2) is - 0.0018 and the adjusted R-squared of the second simple regression model (Figure 3) is 0.0094. The adjusted R-Squared of the first multivariate regression model (Figure 4) is – 0.0028 and the adjusted R-squared of the second multivariate regression mode (Figure 7) is 0.0098. It can be concluded that all values of the adjusted R-squared of the models are very low, since the independent variables of all models explains less than a percent of the variation in the dependent variable. As a result of this outcome, it is not wise to choose one of the four models based on the adjusted R-squared. The models are used to forecast the stock market. But the truth is that these are unpredictable. This explains the low value of the adjusted R-squared of the models.

5.3.2 Information criteria

The information criteria is a function in Stata that makes it possible to compare the performance of different models. These two information criteria are called AIC and BIC. For all models, the smallest value of the AIC or BIC, the better the model performs. The results of the test for all the models are shown below in figure 8 (simple regression model using one month lag), figure 9 (simple regression model using one day lag). Figure 10 (multivariate regression model using one month lag) and figure 11 (multivariate regression model using one day lag).

Figure 8: Information criteria simple regression model using one month lag

Akaike's information criterion and Bayesian information criterion

Model	N	ll(null)	ll(model)	df	AIC	BIC
.	**328**	**573.1754**	**573.3878**	**2**	**-1142.776**	**-1135.19**

Figure 9: Information criteria simple regression model using one day lag

Akaike's information criterion and Bayesian information criterion

Model	N	ll(null)	ll(model)	df	AIC	BIC
.	**6,901**	**20874.21**	**20907.2**	**2**	**-41810.41**	**-41796.73**

Figure 10: Information criteria multivariate regression model using one month lag

Akaike's information criterion and Bayesian information criterion

Model	N	ll(null)	ll(model)	df	AIC	BIC
.	**328**	**573.1754**	**573.722**	**3**	**-1141.444**	**-1130.065**

Figure 11: Information criteria multivariate regression model using one day lag

Akaike's information criterion and Bayesian information criterion

Model	N	ll(null)	ll(model)	df	AIC	BIC
.	**6,901**	**20874.21**	**20909.16**	**3**	**-41812.32**	**-41791.8**

The value of the AIC and BIC of the simple regression model and the multivariate regression model both lagged with one day are the smallest. When we compare these two models based on AIC and BIC, it can be concluded that the simple regression model is the lowest over both values. This confirms that the simple regression model outperforms the multivariate regression model. When we compare the AIC and BIC of the simple regression and multivariate regression model both lagged with a month, we can also conclude that the simple regression model performs better based on the value of the AIC and the BIC.

Based on the adjusted R-squared and the conclusion based on the information criteria, it can be concluded that the simple regression model is a better predictor of returns than the multivariate regression model. This applies to the models based on a one month lag and the models based on a one day lag. Based on this, it can be concluded that the VIX does not contribute to predicting the returns of the S&P 500 index.

5.4 Rule of thumb

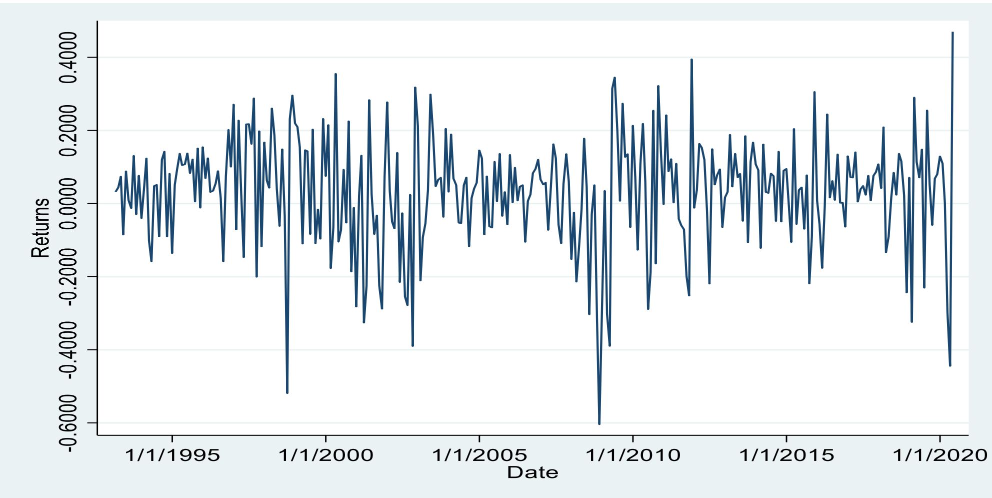
In this paragraph, the rule of thumb is drawn up on the basis of the best performing model. First, returns are calculated in Excel using the model. These values are then plotted and described. Finally, the rule of thumb is drawn up, this rule indicates when the investor must enter the marker and when the investor must exit the market.

As mentioned earlier, the simple regression model performs better than the multivariate regression model. This means that based on my analysis the VIX does not make the multivariate regression model perform better. The model that I use to predict the returns is the simple regression model lagged with one month. Although according to the information criteria in Stata, the model with one day lag performs better I decided to use the model with one month lag. After all, when predictions are made daily, there are many more opportunities to enter or exit the market. I want to create a rule that reduces this number of trades and so I choose to use the simple regression model with one month lag when setting up the rule of thumb. This results in fewer predictions, but based on these predictions I want to set up the rule. The formula I use during the rule of thumb is shown below:

$$Rm_t = 0.00656 + 0.03603 * Rm_{t-1} + u_t \qquad \text{(III)}$$

The first step is to calculate the expected returns of the S&P 500 index based on the simple regression model. During this step the returns are calculated by means of the model based on the monthly returns lagged with 1 month. This operation is performed using Excel (see appendix I). Subsequently, the results of the model are shown below for the graphical representation (see figure 12).

Figure 12: Predicted returns using the simple regression model



This graph shows the returns predicted on the basis of a simple regression model. The returns are predicted based on the one month lag of the historical returns.

Figure 12 shows that the predicted returns vary greatly. According to the summary statistics the lowest predicted return is 60.38 percent negative and the highest predicted return is 47.00 percent positive. In addition, the mean of the 328 observations is 3.07 percent and the standard deviation is 15.19 percent. Based on this information, I will draw up the rule of thumb, which must ensure that the investor is in the market when the model predicts profits. In addition, the investor must be out of the market when negative returns are predicted based on the model.

The rule of thumb is drawn up on the basis of the mean and the standard deviation. The mean of the predicted returns is 3.07 percent and the standard deviation of the predicted returns is 15.19 percent. It is important that the rule occurs several times in the past 27 (1993 until 2020) years. The rule is not practical if it can be applied 1 or 2 times based on the data of the past 27 years. On the other hand, it is also not practical if the investor has to trade very often.

Based on the returns predicted by means of the simple regression model, I determine the following rules regarding investing in the S&P 500 index or exiting the S&P 500 index.

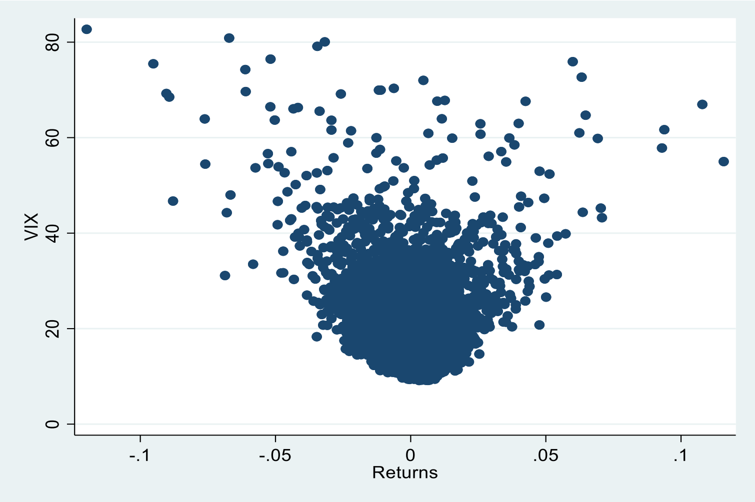
- I will exit the S&P 500 index when the model predicts that the returns will be -12.12 percent or lower. This means I will exit the market when the predicted return is below the mean minus one the standard deviation.
- I will enter the S&P 500 index when the models predicts that return will be 18.26 or higher. This means that I will enter the market when the predicted return is higher than the mean plus one standard deviation.

When the previously described rules are followed by an investor, this investor exits the market 42 times over the past 27 years and enters the market 45 times or buys additional assets at that time in over the past 27 years.

5.5 investment strategy using the VIX

During this last paragraph of the results chapter, an investment strategy, based on the value of the Volatility Index VIX, is tested. First of all a scatter plot is added for the graphical representation of the relationship of the VIX and the returns of the S&P 500 index. Then, it is told how this test is performed. Furthermore, it is described how the test will be carried out and finally the results of testing the strategy are analysed. Based on these results, it is concluded whether the strategy is valuable.

Figure 13: Graphical representation of the relationship between the VIX and the returns



This scatter plot shows the relation between returns of the S&P 500 index and the value of the VIX. The period on which the scatter plot is based runs from January 1993 to May 2020.

Figure 13 shows a scatter plot based on the value of the VIX and the returns of the S&P 500 index. From this scatter plot a relationship between the VIX and the returns of the S&P 500 index can be deduced. The fact is that when the VIX increases, the returns spread more, both positive and negative compared to when the VIX is lower. In addition, it is generally the case that returns tend to turn negative rather than positive when the VIX increases. This conclusion is confirmed by the table in Figure 14. This model is a regression of the returns based on the lagged VIX. According to the model, the returns decrease by 0.168 percentage point when the VIX increases by one unit. In addition, this variable is significant, as the t-statistic is -5.99. This confirms that as the VIX increases, returns tends to decrease.

Figure 14: Simple regression model based on the one month lagged VIX

VARIABLES	(1) Rmt
VIX_Lagged	-0.00168***
	(-5.99)
Constant	0.03944***
	(6.69)
Observations	328
Adjusted R-squared	0.0963

This table shows the result of one regression from the period May 1993 until January 2020. The dependent variable is return of the market (S&P 500 index) and the independent variable is the one month lag of the value of the VIX. The numbers in the parentheses are the t-statistics. *, **, *** indicates the significance levels of 0.10, 0.05 and 0.01.

During the remainder of this section, I want to take advantage of this relationship and establish a strategy that ensures that an investor is not exposed to these extreme returns.

Earlier in this chapter, different models were used to predict market returns. Two of these models included the lagged VIX as a variable, however it was concluded that this model was less valuable as a predictor of market returns than the simple regression models without the lagged VIX. Based on this, it can be concluded that the VIX does not contribute to predicting returns based on my analysis and data used. However, it must be taken into account that it is in fact not possible to predict the market. But beyond that, the predictions based on the model with only the lagged market returns as variable were better than the model including the lagged VIX as variable.

Earlier in the chapter it is concluded that the VIX does not contribute to predicting returns of the S&P 500 index. However I want to test an investment strategy based on past data regarding the VIX to test whether I can use the VIX to set up an investment strategy. This concerns a strategy when an investor leaves the market at certain times and re-enters the market at certain times, in the case of this thesis the market is the S&P 500 index. I am doing this test to conclude whether it is possible to outperform the market by means of a strategy based on the VIX. When it turns out that the returns achieved with the strategy based on the VIX are higher than when an investor always remains in the market, it can be concluded that this strategy outperformed the market and it might be a strategy that is interesting for investors.

The strategy I want to test has already been mentioned in the empirical model in Chapter 4. The strategy is based on the VIX and is as follows. Using this investment strategy the investor is out of the market when the VIX is higher than the value of the mean plus one standard deviation. This means that the investor leaves the market when the value of the VIX is higher than 27.77. When the VIX has a value of 27.77 again, the investor enters the market again. This strategy should prevent the investor from incurring large losses due to the decline in returns of the S&P 500 index.

Testing of the investment strategy is performed by means of Excel and Stata. I start with the full dataset, which is the same as when the investor always remains invested in the market, during the period from January 1993 to May 2020. This leads to 6,902 observations with an average value of the VIX of 19.37 and an average daily return of 0.035 percent which was achieved over this period. These results are shown below in figure 15.

Figure 15: Summary statistics of the *unrestricted* investment strategy from January 1993 until May 2020

Variable	Obs	Mean	Std. Dev.	Min	Max
VIX	6,902	19.377	8.394	9.14	82.69
DailyReturn	6,902	0.00035	.012	-.12	.116

Subsequently, a filter was applied in Excel that ensures that the daily returns are only included when the value of the VIX is lower than 27.77. An indication of this action is given in Appendix II. This results in a dataset with only daily returns achieved when the VIX is lower than 27.77. Based on this dataset, the summary statistics are run in Stata to show the average daily return of the restricted strategy. This test leads to the following results, which are shown in figure 16.

Figure 16: Summary statistics of the *restricted* investment strategy from January 1993 until May 2020

Variable	Obs	Mean	Std. Dev.	Min	Max
VIX	6,076	17.045	4.638	9.14	27.76
DailyReturn	6,076	.0007	.009	-.038	.05

The summary statistics show that using the restricted strategy there are 6,076 observations. This makes sense as the investor is out of the market when the VIX is 27.77 or higher. In addition, the mean of the VIX is also lower, namely 17.05, this is because the high values of the VIX have been removed from this dataset. The most important outcome of this summary statistic is the value of the daily return. This is 0.07 percent, and this daily return is relatively much higher than the 0.035 percent if the investor always remains in the market. This means that the average daily return based on the investment rule is higher than if the investor always remains in the market. On average, a year has about 250 trading days on the exchange, which results in an average annual return of 8.75 percent when the investor always stays in the market from 1993 to May 2020. When using the investment strategy based on the VIX, an average annual return of 17.5 percent is achieved over the period January 1993 to May 2020. This is a very big difference and can yield large profits to investors.

The results indicate that the strategy is certainly interesting for investors. But it is also possible to determine the significance of the results. Stata is used to analyse whether the coefficient for the average daily returns is significant. This is done by regressing the results for the daily returns to a constant. This leads to the following model which is shown in figure 17.

Figure 17: Model consisting of constant factor of daily returns from January 1993 until May 2020 for testing significance.

VARIABLES	(1) DailyReturn
Constant	0.00070*** (6.32)
Observations	6,076
Adjusted R-squared	0.00000

This table shows the result of one regression from the period January 1993 until May 2020. This is a regression of the daily returns on a constant to determine whether the results of the daily return are statistically significant. The numbers in the parentheses are the t-statistics. *, **, *** indicates the significance levels of 0.10, 0.05 and 0.01.

The value of the t-statistic is 6.32 and this results in a p-value is 0.000. This means that the coefficient is significant at 1 percent significance level. This reinforces the strategy based on the VIX. This strategy is also less risky as the variance of the daily returns in the restricted strategy is lower than the variance in the unrestricted strategy. In addition, the standard deviation of the unrestricted is higher than that of the restricted strategy.

Since the period from January 1993 until May 2020 is a very long period and data and times change, I also want to test the strategy based on data from January 2017 until May 2020 to make the results more recent. The data related to this dataset is shown in Figure 18 below.

Figure 18: Summary statistics of the *unrestricted* investment strategy from January 2017 until May 2020

Variable	Obs	Mean	Std. Dev.	Min	Max
VIX	857	16.643	9.544	9.14	82.69
DailyReturn	857	0.00044	.013	-.12	.094

The dataset consists of 857 observations. The mean of the VIX during this period is 16.64, the standard deviation is 9.54, and the mean daily return over this period is 0.044 percent. This means that during this period, the investment strategy is to exit the market when the VIX is 26.18 or higher. 26.18 (16.64 + 9.54) Is the value of the mean of the VIX plus the standard deviation of the VIX during the period from January 2017 until May 2020. Using this value of the VIX the summary statistics are calculated which is shown below in figure 19.

Figure 19: Summary statistics of the *restricted* investment strategy from January 2017 until May 2020

Variable	Obs	Mean	Std. Dev.	Min	Max
VIX	779	14.157	3.576	9.14	25.61
DailyReturn	779	.00057	.008	-.034	.034

The summary statistics show that using the restricted strategy there are 779 observations. The mean of the VIX is 14.15, this is because the high values of the VIX have been removed from this dataset. The most important outcome of this summary statistic is the value of the daily return. This is 0.057 percent, and this daily return is also higher than the 0.044 percent if the investor always remains in the market. This means that the average daily return based on the investment rule is higher than if the investor always remains in the market. On average, a year has about 250 trading days on the exchange, which results in an average annual return of 11 percent when the investor always stays in the market from January 2017 to May 2020. When using the investment strategy based on the VIX, an average annual return of 14.25 percent is achieved over the period January 2017 until May 2020. This means that through the restricted investment strategy, the annual return is 3.25 percentage points higher in the period from January 2017 to May 2020.

The results of this sample also indicate that the strategy is certainly interesting for investors. The significance of the coefficient is also determined for this sample. This leads to the following model which is shown in figure 20.

Figure 20: Model consisting of constant factor of daily returns from January 1993 until May 2020 for testing significance.

VARIABLES	(1) DailyReturn
Constant	0.00058** (2.13)
Observations	779
Adjusted R-squared	0.00000

This table shows the result of one regression from the period January 2017 until May 2020. This is a regression of the daily returns on a constant to determine whether the results of the daily return are statistically significant. The numbers in the parentheses are the t-statistics. *, **, *** indicates the significance levels of 0.10, 0.05 and 0.01.

The value of the t-statistic is 2.13. This means that the coefficient is significant at 5 percent significance level. Although it is not the case that the coefficient is significant at 1 percent. However, this coefficient is significant at 5 percent, which is positive and strengthens the investment strategy. This strategy is also less risky as the variance of the daily returns in the restricted strategy is lower than the variance in the unrestricted strategy. In addition, the standard deviation of the unrestricted is higher than that of the restricted strategy.

Based on the two outcomes in the different samples, it can be concluded, that higher returns can be achieved by means of the investment strategy based on the VIX. In comparison with the unrestricted strategy when the investor always remains in the market. This conclusion is reinforced as the results are also significant. For the period from January 1993 to May 2020, the significance is 1 percent and the period January 2017 to May 2020, the significance is 5 percent.

6. Conclusion

The aim of this thesis was to investigate whether there is a relationship between the returns of the S&P 500 index and the value of the VIX. In addition, during the thesis it was investigated whether the VIX has predictive power and whether the VIX can be used during the development of an investment strategy. These aspects are the guiding principle of the thesis and lead to the following research question:

Is there a relationship between the returns of the S&P 500 index and the VIX, and is it possible to set up a profitable trading strategy for the S&P 500 index using the VIX?

To answer this research question, an analysis is performed during the thesis based on various datasets. These datasets were mainly collected through WRDS, data that was not yet available through WRDS is collected through Yahoo Finance. The dataset used in the thesis is the value of the VIX between January 1993 and May 2020. In addition, the returns of the S&P 500 index are used in this thesis. The datasets regarding the returns of the S&P 500 index are also from the period January 1993 to May 2020.

The empirical strategy started with the creation of the simple regression models and the multivariate regression models. Subsequently, an investment rule of thumb was drawn up based on one of the models. The choice of the model is based on the adjusted R-squared and the information criteria. Thereafter, it was analysed whether a relationship can be discovered in the value of the VIX and the performance of the returns of the S&P 500 index. Finally, an investment strategy is drawn up based on the VIX. This strategy involves exiting the market when the value of the VIX is equal to the mean of the VIX plus the value of a standard deviation. The mean and standard deviation have been calculated over the entire period from January 1993 to May 2020, as well as over the period from January 2017 to May 2020.

The simple regression models and multivariate regression models were first created in the chapter results. Based on the adjusted R-squared and the information criteria, I concluded that the simple regression model performs better than the multivariate regression model. This applies to the models based on a one month lag and the models based on a one day lag. The multivariate regression model included the lagged VIX. Based on the choice of the model and the data I used, it can be concluded that the VIX does not contribute to predicting returns of the S&P 500 index.

Subsequently, a rule of thumb was drawn up based on the predicted returns based on the simple regression model. Based on the returns predicted by means of the simple regression model, I determine the following rules regarding investing in the S&P 500 index or exiting the S&P 500 index.

- I will exit the S&P 500 index when the model predicts that the returns will be -12.12 percent or lower. This means I will exit the market when the predicted return is below the mean minus one the standard deviation.
- I will enter the S&P 500 index when the models predicts that return will be 18.26 or higher. This means that I will enter the market when the predicted return is higher than the mean plus one standard deviation.

Thereafter, the investment strategy based on the VIX was tested. This strategy was tested on the basis of two different periods. The entire period that the VIX exists, from January 1993 to May 2020 and the period from January 2017 to May 2020. When the investment rule is applied to the entire sample from January 1993 to May 2020. The average daily return using the investment rule is 0.07 percent, this daily return is relatively much higher than the 0.035 percent if the investor always remains in the market. A year has about 250 trading days on the exchange, which results in an average annual return of 8.75 percent when the investor always stays in the market from 1993 to May 2020. When using the investment strategy based on the VIX, an average annual return of 17.50 percent is achieved. This is a very big difference and can yield large profits to investors.

For the period from January 2017 to May 2020, the average annual return on the returns achieved through the investment strategy is also higher. Using the investment strategy the average annual return is 14.25 percent, in comparison with 11 percent when the investor always stays in the market. This means that through the restricted investment strategy, the annual return is 3 percentage points higher in the period from January 2017 to May 2020. Based on the two outcomes in the different samples, it can be concluded, that higher returns can be achieved by means of the investment strategy based on the VIX. This conclusion is reinforced as the results are also significant. For the period from January 1993 to May 2020, the significance is 1 percentage and for the period January 2017 to May 2020, the significance is 5 percentage.

Based on the results, it can be concluded that there is a relationship between the VIX and the returns of the S&P 500 index. When the VIX is high, the returns from the S&P 500 index are extremely widespread and these returns tend to decrease. In addition, the VIX can be used while investing in the S&P 500 index, as the investment strategy based on the VIX performs better than if the investor always stays in the market.

Finally, there are a number of limitations to this research. In this thesis, an attempt was made to predict market returns based on historical returns and the VIX. Predicting the market is actually not possible, which is why the R-squared of both models is also so low. It is true that one of these two models was used to draw up the rule of thumb in chapter 5.4. In addition, an investment strategy based on the VIX is drawn up in chapter 5.5. This strategy performs better than when the investor stays in the market in both tested samples. However, during the analysis of these results in chapter 5.5, It does not take into account transaction costs, which arises when the investor buys and sells ETFs.

Bibliography

Ang et al. . (2009). *High idiosyncratic volatility and low returns: International and further U.S. evidence.* Journal of Financial Economics.

Baillie, R. T., & DeGennaro, R. P. (1990). *Stock returns and volatilty.* Journal of Financial and Quantitative analysis.

Barro, R. J., & Ursuá, J. F. (2017). *Stock-market crashes and depressions.* Research in Economics.

Becker, R., Clements, A., & McClelland, A. (2009). *The jump component of S&P 500 volatility and the VIX index.* Journal of Banking & Finance.

Bekaert, G., & Hoerova, M. (2014). *The VIX, the variance premium and stock market volatility.* Journal of Econometrics.

Blair, B., Poon, S.-h., & Taylor, S. (2001). *Forecasting S&P 100 volatility: the incremental information content of implied volatilities and high-frequency index returns.* Journal of Econometrics .

Bushee, B. J., & Goodman, T. H. (2007). *Which Institutional Investors Trade Based on Private Information About Earnings and Returns?* Journal of Accounting Research.

Chang, C.-C., Hsieh, P.-F., & Lai, H.-N. (2009). *Do informed option investors predict stock returns? Evidence from the Taiwan stock exchange.* Journal of Banking & Finance.

D.Gibson, H., G.Hall, S., & S.Tavlas, G. (2012). *The Greek financial crisis: Growing imbalances and sovereign spreads.* Journal of International Money and Finance.

Daube, C. H. (2020 (working paper)). *The Corona Virus Stock Exchange Crash.* ZBW.

Ederington, L. H., & Guan, W. (2002). *Is implied volatility an onformationally efficient and prector of future volatility?* Journal of Risk.

Fernandes, M., Medeiros, M., & Scharth, M. (2014). *Modeling and predicting the CBOE market volatility index.* Journal of Banking & Finance.

French, K. R., & Schwert, G. W. (1987). *Expected Stock Returns and Volatility .* Journal of Financial Economics.

Goonight, G. T., & Green, S. (2010). *Rhetoric, Risk, and Markets: The Dot-Com Bubble.* Quarterly Journal of Speech.

Guo, H., & Savickas, R. (2012). *Idiosyncratic Volatility, Stock Market Volatility, and Expected Stock Returns.* Journal of Business & Economic Statistis.

Kownatzki, C. (2016). *How Good is the VIX as a Predictor of Market Risk? .* Journal of Accounting and Finance.

Marquering, W., & Verbeek, M. (2004). *The Economic Value of Predicting Stock Index Returns and Volatility.* Journal of Financial and Quantitative Analysis.

Mishkin, F. S. (2011). *Crisis, Over the Cliff: From the Subprime to the Global Financial .* Journal of Economic Perspectives.

Sarwar, G. (2012). *Is VIX an investor fear gauge in BRIC equity markets?* Journal of multinational Finance.

Shaikh, I., & Padhi, P. (2015). *The implied volatility index: Is 'investor fear gauge' or 'forward-looking'?* Borsa Istanbul Review.

Szakmary, A., Ors, E., Kim, J. K., & Davidson, W. N. (2003). *The predictive power of implied volatility: Evidence from 35 futures markets.* Journal of Banking & Finance.

Tsai, W.-C., Chiu, Y.-T., & Wang, Y.-H. (2015). *The Information Content of Trading Activity and Quote Changes: Evidence from VIX Options.* Journal of Futures Markets.

Wang, J.-C. (2008). *Investigating market value and intellectual capital for S&P 500.* Journal of Intellectual capital.

Whaley, R. E. (2000). *The investor Fear Gauge.* The journal of portfolio management.

WRDS. (2020). *About*. Retrieved from WRDS: https://wrds-www.wharton.upenn.edu/pages/about/

Yahoo Finance. (2020). *Finance*. Retrieved from Finance.Yahoo: https://finance.yahoo.com/

Appendix I

Date	Rm_{t-1}	Rm_t using the model
26-2-1993	0,70%	3,19%
31-3-1993	1,05%	4,43%
30-4-1993	1,87%	7,39%
28-5-1993	-2,54%	-8,50%
30-6-1993	2,27%	8,84%
30-7-1993	0,08%	0,93%
31-8-1993	-0,53%	-1,26%
30-9-1993	3,44%	13,06%
29-10-1993	-1,00%	-2,94%
30-11-1993	1,94%	7,64%
31-12-1993	-1,29%	-4,00%
31-1-1994	1,01%	4,29%

Appendix II

Date	VIX	Daily Return
11-3-1998	27,76	-0,068%
5-11-2000	27,76	1,790%
12-11-2002	27,76	0,056%
8-22-2002	27,75	1,405%
3-28-2003	27,75	-0,578%
2-7-2018	27,73	-0,500%
12-11-1998	27,72	0,124%
11-4-2009	27,72	0,104%
11-22-2000	27,71	-1,855%
12-18-2000	27,7	0,807%
8-14-2007	27,68	-1,816%
2-1-1999	27,67	-0,519%
8-9-1999	27,66	-0,192%