How Well Does Implied Volatility Predict Future Stock Index Returns and Volatility?

A study of Option-Implied Volatility Derived from OMXS30 Index Options

Authors: Sara Vikberg and Julia Björkman

Stockholm Business School
Bachelor's Degree Thesis 15 HE Credits
Subject: Business Administration
Spring Semester 2020
Supervisor: Jarkko Peltomäki

Stockholm Business School

Stockholm University

Abstract

The purpose of this thesis is to study if and how well implied volatility can predict realised volatility and returns on the OMXS30 index one month in the future. The findings are put in relation to how historical volatility can predict realised volatility and how changes in implied volatility can predict returns. The study covers the time period from 10th of May 2012 to 9th of February 2020 and the implied volatility used in the study is derived from an unweighted average of OMXS30 call and put option implied volatility. Six different OLS-regressions are performed to study the prediction capability of implied volatility. This study finds support of implied volatility to be a statistically significant estimate for future realised returns in a univariate regression. However, our results show that historical volatility performs slightly better predictions of realised volatility than implied volatility. These are contradictory results to the majority of the papers studied in this thesis. These papers share the common notion that implied volatility is superior to historical volatility in predicting realised volatility. Further our results show that implied volatility nor change in implied volatility are significant estimates to future realised returns and perform poorly as predictors. This result is supported by the larger part of previous research, which found implied volatility to be a weak predictor of returns.

Note: Realised volatility in this essay is referred to as volatility on period ahead while historical volatility is referred to the volatility on period back

Table of contents

1. Introduction

1.1 Background

The possibility to determine the future value of financial instruments has been widely researched through the years and many theories and models have been developed in order to predict the future of the stock market. These developed theories and models have been used to create different forecasting methods. The forecasting methods used to attempt to predict the stock market can be divided into two types, fundamental and technical analysis (Gunn, 2009). Fundamental analysis attempts to predict future movements of the stock market by measuring the intrinsic value of a financial instrument by evaluating aspects of businesses or markets. Examples of aspects evaluated in fundamental analysis are microeconomic factors like revenue and company management or macroeconomic factors such as the state of the economy. A value is assigned to the financial instrument after the fundamental analysis evaluation to determine whether it is under- or overvalued on the market (ibid.).

Fundamental analysis is based on the weak form of the efficient-market hypothesis, developed by Eugene Fama in 1970. This means that the security price of today reflects all the data of past prices and no form of historical data analysis is effective for investment decisions. In contradiction, technical analysis does not confirm the efficient market hypothesis and is based on historical technical data. Technical analysis aims to identify investment opportunities by analysing statistical trends. These trends can be studied on data such as the historical trading volume or price of a security (Gunn, 2009). Fundamental analysis is based on the work of Dow Jones Index founder, Charles Dow, and his two basic assumptions. These two assumptions are that markets are efficient, but that even random market movements seem to move in identifiable patterns that tend to repeat over time (ibid.).

Options contracts is a type of financial derivative that contains information that can possibly be used together with technical analysis to predict upcoming market movements. Over the last two decades the use of derivatives has increased dramatically around the world (Hull, 2017). The new technology of this time has made online trading possible and made many different financial instruments accessible for the public (Stowell, 2012). In the modern options market thousands of contracts are listed on exchanges and millions of contracts are traded daily (ibid.). Options markets are much more well-functioning and efficient than 20 years ago and the options prices reflect the market view of many more investors (ibid.). These views of trading investors are reflected in the prices of options and is information that could possibly be used to forecast the price of the underlying asset of the option (Hull, 2017).

Implied volatility is a form of derivative information that can be extracted from option prices. Through the Black-Scholes option pricing model it is possible to derive the volatility that is implied by the option price, known as the implied volatility. This implied volatility is a forward-looking measure that can be interpreted as the market's expectations of the future volatility of the option's underlying asset (Hull, 2017). The level of implied volatility extracted from options on stock indexes can furthermore be an indicator of market fear and future instability. A high level of implied volatility indicates great uncertainty on the market and a lower level of implied volatility indicates a more stable future market (ibid.).

The link between investor information and implied volatility can be understood under demand-based option pricing theory. If option traders possess positive (negative) information resulting in a bullish (bearish) view of the future stock market, they will either buy (sell) call options or sell (buy) put options. There will be limited capacity for market markers to meet customer demand. Hence, the demand pressures associated with positive (negative) information translates into price pressures, which pushes implied volatility higher (Bing & Gang, 2017). As the expectation of future volatility changes or as the option demand increases, implied volatility will rise and result in high-priced option premiums (Nations, 2012).

1.2 Problematization

There is a large and growing literature documenting the prediction ability of information extracted from the equity options market. Despite the fact that equity options are sensitive to macroeconomic factors, there are conflicting views on whether equity options contain useful information to predict future stock movements and whether this information is already incorporated in the stock market (Duan & Wei, 2008). The future stock movements that we will be focusing on in this study is realised volatility and returns.

Implied volatility is known to covary with realised volatility and is generally claimed to be superior to past realised volatility in predicting future volatility (Christensen & Prabhala, 1998; Christensen & Hansen, 2002; Szakmary et al, 2003). Whether implied volatility predicts realised future volatility better than past realised volatility has been widely tested in research papers, especially during the 1980's and 1990's. These research papers have reached mixed results and conclusions.

Day & Lewis (1992) found that implied volatilities perform equally to forecasts from past realised volatility. Contrarily, Canina & Figelwski (1993) found conflicting evidence to Day & Lewis. Their study found that simple past realised volatilities outperform the poor volatility forecasts of implied volatilities. Christensen & Prabhala (1998) provided evidence that implied volatilities are unbiased forecasts of volatility and that they outperform historical information models in forecasting volatility.

Additionally, there has been extensive research regarding how implied volatility can predict realised returns. Hence, if there is a relationship between implied volatility and returns. The main finding from previous research is that there is a negative as well as an asymmetric relationship between the two variables (Bekiros et al (2017); Thakolsri, Sethapramote & Jiranyakul, 2016). The conclusion regarding the prediction power of implied volatility is however not as unanimous. Egbers & Swinkels (2015) and Dennis et al (2006) found change in implied volatility to be a weak predictor for returns whereas Amman, Verhofen & Süss (2008) and Rubbaniy et al (2014) concluded the opposite.

1.3 Aim and Contribution

Compared to other studies in this area we will focus this study entirely on short term relations between the implied volatility from OMXS30 index options and future 1) realised volatility and 2) returns on the OMXS30 index. The aim of this study will therefore be to analyse how information in option prices, revealed in the form of implied volatility, is related to upcoming short-term volatility and returns on the OMXS30 index. The mixed results from previous research regarding the forecasting effectiveness of implied volatility encouraged us to test this and see if findings from previous studies hold on a more recent data set and a non-American market. Most previous studies regarding implied volatility we found are based on an American market. We want to examine if implied volatility derived from OMXS30 index options can predict monthly realised volatility and returns on the OMXS30 index. The data set chosen to examine this is for the period May 2012 to February 2020. This is due to the fact that the implied volatility data we could access from Thomson Reuters Datastream on OMXS30 options was available from May 2012.

Understanding whether and how the stock market can be predicted is important for portfolio allocations and market efficiency. We would like to contribute to this understanding of stock market prediction of previous research with our study. The market valuation of stocks tends to be evidently unpredictable and stock prices can increase or decrease without known reasons. Therefore, investors would benefit from having a supplement, like implied volatility levels, to their analysis methods in order to increase the probability of profitable stock market investments. If implied volatility were to be an efficient predictor of realised volatility and future returns it could serve as an analysis source for investors and speculators.

The contribution to previous research is that it differs from most previous studies in that it is European market based and based on more recent data. Using more recent data is of interest since previous research is conducted mostly around the 1990's and there has been vast technical development since then. It is of interest to examine whether a study based on the modern and more technically developed stock and derivatives market of today, with for example larger trading volumes and trading robots, yields different results to the studies based on past data of a different decade (Thomson Reuters Datastream). Additionally, this study differs from some previous studies in that it will only study short-term stock market predictions of one month. To minimise the risk of overlapping volatility data in our regression and keep our study statistically robust we have chosen to examine non-overlapping monthly periods. We chose this short-term prediction period and non-overlapping data since some previous studies, as Christensen & Prabhala (1998) and

Christensen & Hansen (2002), used monthly non-overlapping data to avoid issues with autocorrelation in regressions.

The two main research questions of this study are thus the following:

1. If and how well does implied volatility derived from OMXS30 options predict *realised index volatility*?

2. If and how well does implied volatility derived from OMXS30 options predict *realised index returns*?

1.4 Research Approach

The research approach conducted in this study is quantitative and deductive. The study is limited to the OMXS30 index and OMXS30 index options on the Nasdaq Stockholm Stock Exchange. Furthermore, only short-term relations of one month will be studied. The data sample consists of monthly implied volatility of OMXS30 index options and returns on the OMXS30 index between the 10th of May 2012 and 9th of February 2020. The returns are used to calculate the past historical volatility and realised volatility and the investigation of the hypotheses will be made through the method of Ordinary Least Square (OLS) regression in Eviews. Four separate regressions will be conducted where the aim is to find out whether and how well implied volatility predicts future realised volatility and returns.

2. Literature Review

This chapter begins with an introduction to general option theory and the Black-Scholes-Merton model for option pricing which is then followed by theory about volatility, historical volatility and implied volatility. The literature review is then concluded with a literature survey of previous research.

2.1 Theoretical Framework

2.1.1 Option Theory

Options contracts are a type of financial derivative, which are contracts between a buyer and a seller, and are mainly divided into two different types: call and put options (Hull, 2017). A call (put) option gives the buyer the right to buy (sell) the shares of an underlying asset at a predetermined price, the strike price, within a specified time limited by a specific date, the maturity date (ibid.). Options contracts differ from other derivatives, such as futures, in that they give the option holder the right, but not the obligation, to exercise the contract and that the investor must pay an option premium for the options contract whilst it costs nothing to enter into a futures contract (ibid.).

Furthermore, options can be divided into two additional categories, European and American options, which differ in regard to when they can be exercised. European options can only be exercised at the maturity date, whereas American options can be exercised at any date up until the maturity date (Hull, 2017). Stock index options are more commonly of European style, while stock options are more commonly American style.

Options are often referred to as being either in-the-money, at-the-money, near-the-money or out-of-the-money (Hull, 2017). For call options, if the stock price is greater than the strike price ($S > K$) the option is said to be in-the-money and if the stock price is less than the strike price ($S < K$) the option is said to be out-of-the-money and for put options it is thus the opposite. When the stock price of an option is the same as the strike price ($S = K$) the option is said to be at-the-money. A call option where the strike price is lower than the stock price, but extremely close to it, is considered to be near-the-money and put options are near-the-money when the strike price is higher than the stock price, but extremely close to it. Options will only be exercised when it is in the money or near the money, as that is when one makes a profit (ibid.).

There are six factors according to option theory that affects the option price (Hull, 2017):

1. The current stock price, S_0
2. The strike price, K
3. The time to expiration, T
4. The volatility of the stock price, σ
5. The risk-free interest rate, r
6. The dividends expected during the life of the option

2.1.2 The Black-Scholes-Merton Model

In the early 1970's, the Nobel Prize awarded Black-Scholes-Merton model was introduced to the world by founders Fisher Black, Myron Scholes and Robert Merton (Hull, 2017). The Black-Scholes-Merton model, also known as the Black-Scholes model, is the first widely used model for option pricing and is still today one of the most important concepts in modern financial theory (ibid.**)**. This model has had a significant influence on financial engineering and how traders price and hedge options (ibid.).

The model calculates the value of a European option using inputs that are all observable in the market, with one exception being volatility. The stock price is known, the risk-free rate is usually set equal to the risk-free rate of an investment that lasts for time T and the option price as well as time to maturity are specified in the options contracts (Hull, 2017). The input volatility is the volatility of the underlying asset during the remaining life of the option and is an ex-post variable, meaning it is calculated on past stock returns. Since current volatility is not observable, a record of historical stock price movements can instead be used to estimate volatility (Hull, 2017). This historical stock price is usually observed at fixed intervals of time, for example during a day, week, or month (ibid.)

The Black-Scholes-Merton pricing formula for European call and put options respectively is:

$$c = S_0 N(d_1) - Ke^{-rT} N(d_2) \tag{1}$$

$$p = Ke^{-rT} N(-d_2) - S_0 N(-d_1) \tag{2}$$

Where:

$$d_1 = \frac{\ln (S_0/K) + (r + \sigma^2/2)T}{\sigma\sqrt{T}} \tag{3}$$

$$d_2 = \frac{\ln (S_0/K) + (r - \sigma^2/2)T}{\sigma\sqrt{T}} = d_1 - \sigma\sqrt{T} \tag{4}$$

Where c is the price of a European call option and p is the price of a European put option (Hull, 2017, p. 325). S_0,K, T, r and σhave already been defined in the "Option theory" section where σ is assumed to be constant. N(x) is the cumulative probability function for a standardized normal variable which gives the probability that a variable with a standard normal distribution will be less than x (Hull, 2017).

Under the Black-Scholes-Merton (BSM) model stock prices are assumed to be normally distributed for short periods of time (Hull, 2017). However, when considering stock prices for longer periods of time, the BSM model assumes stock prices to follow a lognormal distribution (Hull, 2017). This means that the natural logarithm of stock prices, lnS_T, for longer time periods are normally distributed. In contrast to the normal distribution, which is symmetrical around the mean, the lognormal distribution is skewed by the mean. A normally distributed variable can also take on negative as well as positive values whereas a variable with a lognormal distribution is restricted to being only positive (ibid.). Because of this property described it is hence concluded that stock prices follow a lognormal distribution as stock prices cannot be negative (ibid.).

A second assumption of the BSM model is that there are no riskless arbitrage opportunities (Hull, 2017). Further assumptions for the BSM model are that there are no transaction costs or taxes, the trading of a security is continuous, securities can be perfectly divided, and the underlying asset does not pay any dividends during the life of the option (ibid.). Additionally, it is assumed that investors can borrow or lend at the same risk-free rate and that the short-term risk-free rate of interest is constant (ibid.).

2.1.3 Volatility

Volatility, denoted as σ, of a stock price is both a measure of the uncertainty of future stock movements and a measure of the deviation from the mean return (Hull, 2017). The present volatility level is not directly available in the market since it is an ex-post measurement, meaning it can only be calculated based on historical returns (ibid). Therefore, in models such as the Black-Scholes-Merton model it is necessary to predict the volatility parameter with an appropriate method. This prediction can be done through historical volatility or implied volatility.

2.1.4 Historical Volatility and Realised Volatility

Historical volatility, also known as realised volatility, measures the volatility of an asset for a past predefined time interval. This time period for volatility measurement can range from intraday to daily, monthly or yearly (Hull, 2017). Historical volatility can be used in the Black-Scholes-Merton option pricing model as the volatility parameter. In the BSM model volatility is assumed to be constant but in reality, it is changing over time (ibid.). For this thesis historical volatility is referred to as the past realised volatility and realised volatility is referred to as future realised volatility. Even though we refer to them as different in this thesis they are calculated with the same formula below.

The formula for annualised historical volatility and realised volatility is:

$$\sigma = \sqrt{\frac{\sum (r_t - r_m)^2}{N - 1}} \times \sqrt{252} \tag{5}$$

Where:

$$r_{\mathrm{m}} = \frac{\sum \ln\left(\frac{S_t}{S_{t-1}}\right)}{N} \tag{6}$$

$$r_t = \ln\left(\frac{S_t}{S_{t-1}}\right) \tag{7}$$

Where σ, N, r_m, r_t and S_t is the historical volatility, number of observations, mean return, realised return at time t and stock price at time t respectively (Hull, 2017, p. 320, p. 457). The number 252 represents the trading days in one year and it's square root is used to make daily volatility annualised. This is because implied volatility is expressed in annualised terms and therefore we need to annualise the historical volatility as well. The index returns r_t will in this essay henceforth refer to realised returns that are the profit made or lost on an investment over a period of time.

2.1.5 Implied Volatility

Implied volatility, denoted as σ_{IV}, is an important variable in the Black-Scholes-Merton model and is estimated by inverting the call option formula as the following:

$$\sigma_{IV} = BS^{-1}(S_0, t, K, r, \sigma) \tag{8}$$

Where:

$$BS(S_0, t, K, r, \sigma) = S_0 N(d_1) - Ke^{-rT} N(d_2) \tag{9}$$

Where BS stands for the Black-Scholes-Merton model, $\sigma_{IV}$is implied volatility and S_0,K, T, r, σ, N(d_1) and N(d_2) have already been defined in the "Option Theory" section.

Implied volatility, expressed as an annualised percentage (Hull, 2017), is a measure of the market's view of the likelihood that changes of a given security's price will occur. Hence, it is a market's forecast of likely movements of a security's price (Zhang & Nault, 2019). Since future volatility cannot be measured, investors have to make a fair prediction of current volatility instead. Calculating implied volatility from the options market is one method of predicting future volatility. For example, investors can use implied volatility to estimate the future volatility of asset prices based on certain predictive factors, such as option supply and demand and the date to maturity (ibid.). Additionally, investors can value and hedge options with implied volatility information.

Whereas historical volatility is backward looking, implied volatility is forward looking. It is therefore not surprising that predictions based on implied volatility for future asset volatility are known to be slightly superior to predictions based on historical volatility (Hull, 2017). However, while implied volatility can be used to forecast movements in certain securities' prices, it does not predict the direction of this movement. If one observes a high implied volatility, which indicates a large likely move in the underlying asset's price, it is not given whether this movement is strongly upwards or downwards. Hence, implied volatility can give guidance about the underlying asset's future volatility but not the movement's direction (ibid.).

2.2 Literature Survey

2.2.1 Implied Volatility as a Predictor of Realised Volatility

If implied volatility predicts realised future volatility better than historical volatility has been widely tested in research papers, especially during the 1980's and 1990's. These research papers have produced mixed results and conclusions. Most of these research papers have concluded that implied volatility covaries with realised volatility and is generally claimed to be superior to historical volatility in predicting realised volatility. The larger part of these studies has been made on a highly liquid options market, mainly on American S&P 100 index options.

Day and Lewis (1993) studied the predictability of weekly implied volatility derived from near-the-money S&P 100 index call options. The sample data consisted of closing prices and contract volumes for S&P 100 call options and daily closing prices of the underlying S&P 100 index. The sample period stretched from March 1983 to December 1989. The paper concluded that weekly implied volatility contains information about future volatility, but that it performs equally to forecasts from historical volatility.

Canina and Figelwski (1993) studied daily implied volatility of S&P 100 at-the-money, in-the-money and out-of-the-money call options for different subsamples for the sample period March 1983 to March 1987. The paper found that implied volatility yielded poor volatility forecasts and that simple historical volatility outperformed implied volatility when predicting realised volatility. They concluded that implied volatility has virtually no correlation with realised volatility and that it does not incorporate the information in recent historical volatility. In conclusion, they found implied volatility to be a biased and inefficient estimator of realised volatility. The paper found that at-the-money and near to expiration options provides more useful implied volatilities for predicting realised volatility.

Christensen and Prabhala (1998) examined monthly implied volatilities of at-the-money call options on the S&P 100 index over the period November 1983 to May 1995. They used non-overlapping monthly observations and options with about 24 days to maturity. Non-overlapping meaning that each option in the data set expires before the next option is sampled and that each observation is based on its own unique window. The study took the natural logarithm of the volatility data due to its better finite sample properties. This study provided evidence that implied volatility is an unbiased and efficient estimator of realised volatility and that they outperform historical information models in forecasting volatility. Additionally, they found that mixtures of the two forecasts in a multiple regression outperform them both individually. According to the authors their results differ from previous studies due to that they use a longer period series data and use non-overlapping data. Further, the authors argue that overlapping samples leads to problems in statistical analysis which can, such as in Canina and Figelwski (1993) and Day and

Lewis (1993) according to the authors, lead to strong autocorrelation and possible errors in regressions.

Christensen and Hansen (2002) studied the relationship between realised volatility and the trade weighted implied volatility from both in-the-money and out-of-the-money S&P 100 call and put options. The sample data was collected during the period April 1993 to 1997. The study used stock index data and implied volatility data collected from the Thomson Reuters Datastream and the natural logarithm has taken on the volatility data. This paper confirmed the results of Christensen & Prabhala (1998) and confirmed the unbiasedness and efficiency of implied volatility as an estimator of realised volatility. The study found implied volatility, both from call and put options, to be superior to historical volatility in predicting realised volatility. Additionally, the authors found that implied volatility from put options on average is slightly larger than implied volatility from call options. Furthermore, their regression results showed that implied volatility from call options is a better forecast than put implied volatility. Finally, the study concluded that the optimal forecast does not appear to be a trade-weighted average, considering that put options are traded almost as frequently as call options.

Szakmary et al (2003) examined data from 35 futures options markets and eight separate exchanges to test how well implied volatility predicts future realised volatility on the underlying futures. This paper used implied volatility derived from both at-the-money call and put options as an unweighted average. This due to that the implied volatility of call and put options are almost identical. At-the-money options are used since the authors state that implied volatility computed from at-the-money options are least affected by the non-normal distribution of returns and are better predictors of future realised volatility than implied volatility of deep in- or out-of-the-money options. The study uses daily futures data and the implied volatility and realised volatility are measured from 10 to approximately 70 trading days. The historical volatility is calculated using a 30 day window. The regressions are estimated using overlapping observations. Correction of the standard errors of regressions coefficients is done to properly reflect serial correlation of varying lengths in the residuals. The conclusion of the paper is that for a large majority of the futures studied implied volatility outperform historical volatility in predicting realised volatility. Additionally, they found that historical volatility contains no significant predictive information beyond what is already incorporated in implied volatility.

Poon and Granger (2003) compared the research results of 93 different studies about volatility forecasting methods. Among these studies, 34 of them compared the volatility estimate of implied volatility and historical volatility. Out of these 34 studies, 76% or 26 studies concluded that implied volatility provided a better estimate of realised volatility than historical volatility. The remaining 24% or eight studies out of the 34 studies concluded that historical volatility was superior to implied volatility in predicting realised volatility. In this study historical volatility included random walks, absolute returns, historical averages of squared returns, time series models using averages etc. The implied volatility variable included implied volatility based on the Black-Scholes-Merton model and various generalisations. Poon and Granger could conclude from their study that a large majority of the papers considered implied volatility to be a superior predictor over historical volatility when forecasting realised volatility.

2.2.2 Implied Volatility as a Predictor of Realised Returns

The relationship between implied volatility and returns is a widely discussed and researched topic with numerous studies conducted on different markets such as the Thai and American market. Whilst the majority of the studies performed appear to reach similar results that implied volatility may not be an appropriate predictor of returns, there are however studies concluding otherwise.

Bae, Kim and Nelson (2006) conducted a study on whether stock returns and volatility are negatively correlated using a GARCH-model and reached the conclusion that there is a negative correlation between stock returns and volatility. Furthermore, Bekiros et al (2017) studied the relationship between implied volatility and returns by performing a quantile regression. Bekiros et al performed the study on several advanced markets such as the South African, European, Asian and Latin-American markets and covered a time period of 14 years, from 2000 to 2014. Similarly to Bae, Kim and Nelson they discovered an asymmetric and reverse-return relationship.

Thakolsri, Sethapramote and Jiranyakul (2016) conducted a study where they attempted to investigate the relationship between returns and the change in implied volatility on the Thai market. The authors performed an OLS-regression, and they observed an asymmetric relationship as well as a significant negative relationship, which is in line with the results of Bae, Kim and Nelson as well as Bekiros et al's study. Egbers and Swinkels (2015), who studied how well the change in implied volatility performs as a predictor of returns, discovered evidence of implied volatility being a strong predictor of stock returns in a short-term perspective, but a weak long-term predictor. Egbers and Swinkels studied this relationship on the U.S. market with an examination period of 18 years, from 1996 to 2014, using an OLS-regression and modeled implied volatility as a log-linear series. Additionally, contrarily to other previously mentioned studies they did not use the OLS to obtain the p-values associated with the null hypothesis but instead employed a bootstrap method which bootstraps residuals from the regression model.

According to Dennis et al (2006) the asymmetric relationship between implied volatility and returns is a market phenomenon and where this relationship essentially means that negative return shocks tend to imply a future increased implied volatility compared to positive return shocks of the same magnitude. Dennis et al performed this study on S&P 100 index over 7 years, from 1988 to 1995, using an OLS-regression with the change in implied volatility as explanatory variable. To calculate the returns used in the regression they took the natural logarithm of the ratio between daily closing prices, which is in line with the theory of the BSM-model. They reached, as mentioned, the conclusion that there is a negative as well as asymmetric relationship between implied volatility and returns, implying that the change in implied volatility is a weak predictor of returns.

In great contrast to these studies, Amman, Verhofen and Süss (2008) achieved results deviating from the conclusions of an asymmetric and negative relationship. The authors investigated whether implied volatility can predict stock returns on the U.S. stock market using an OLS-regression model on a period of 9 years, from 1996 to 2005. They discovered evidence that implied volatility can be used as a predictor of returns. The authors additionally discovered that there is a positive relationship between returns and lagged implied volatility and that the results appear to be persistent for different times to maturity of implied volatility. In addition to this study, Rubbaniy et al (2014) also found evidence of a positive relationship between implied volatility and returns and that implied volatility is a viable predictor of future returns. The authors performed an OLS-regression to study the U.S. market using the VIX-index to model implied volatility over a period of 19 years, from 1986 to 2005.

2.2.3 Conclusion

In summary, there has been contradictory findings of the informational content and relative prediction qualities of implied volatility. The majority of the broad literature studied seems however to reach similar conclusions. Christensen and Prabhala (1998), Christensen and Hansen (2002) and Szakmary et al (2003) all reached the conclusion that implied volatility is a suitable predictor of future realised volatility and is superior to historical volatility as a predictor. Additionally, out of the 34 papers studied by Poon and Granger (2003) 76% of them concluded that implied volatility was superior to historical volatility in predicting realised volatility. Contrarily, the earlier studies of Day and Lewis (1993) and Canina and Figelwski (1993) concluded that historical volatility is a better predictor of future realised volatility compared to implied volatility. According to Christensen and Prabhala (1998) the conflicting results of Day and Lewis (1993) and Canina and Figelwski (1993) can be due to the issue of overlapping data used in these studies which can lead to autocorrelation and possible regression errors.

The vast majority of the papers studied regarding implied volatility as a predictor of realised returns seems to have reached a unanimous conclusion as well. The majority of the papers studied, including Bekiros et al (2017), Egbers and Swinkels (2015), Thakolsri, Sethapramote and Jiranyakul (2016) and Dennis et al (2006), all concluded that there is a negative asymmetric relationship between returns and implied volatility and/or change in implied volatility. Furthermore, Egbers and Swinkels and Dennis et al concluded that implied volatility is a strong predictor of short-term stock returns but a weak long-term predictor. Despite these unanimous conclusions there are contradictory results, such as one study made by Amman, Verhofen and Süss (2008) as well as one study made by Rubbaniy (2014). These two studies have both reached contradictory conclusions stating that implied volatility can be used as a predictor of returns and that there exists a positive relationship between the variables.

Following the contradictory results from the papers discussed in the previous two paragraphs, it is motivated to conduct further research about implied volatility as a predictor of realised volatility and returns. It is of interest to investigate whether our study confirms the conclusion of the majority of the papers or if we discover contradictory results. The papers studied, specifically Christensen & Prabhala (1998) and Christensen and Hansen (2002), motivates us to use monthly non-overlapping data in our study to avoid possible regression errors caused by autocorrelation. Christensen and Hansen (2002) and Szakmary et al (2003) motivated us to use both implied volatility from call and put options and create an average since, implied volatility of call and put options are almost identical. Using at-the-money options is motivated in several papers, such as Canina and Figelwski (1993), Christensen & Prabhala (1998) and Szakmary et al (2003), since implied volatility from at the money options is claimed to be least affected by the non-normal distribution of returns and act as a better predictor over implied volatility of deep in or out of the money options. Finally, the papers studied, in particular Christensen and Prabhala (1998), Christensen and Hansen (2002) and Egbers and Swinkels (2015), and impelled us to use logarithmical volatility data due to its better finite sample properties compared to non-logarithmized data.

3. Research Design

3.1 Problem, Purpose and Contribution

The purpose of this thesis is to investigate if and how well option-implied volatility predicts future realised volatility and realised index returns. The intention is to investigate how the information reflected in option prices, in the form of implied volatility, is related to short-term future volatility and returns. The studies discussed in the previous section conducted their research using mainly an American market-based sample and a period from before and around the early 2000's. The aim of this thesis is to contribute to this previous research by examining both a non-American market as well as analysing more recent data. Considering that our study concerns a time period that has not been heavily covered by the field of research we consider it to be of great value to examine these relationships under these new conditions. Our study may additionally provide useful results for investors attempting to analyse stock market movements through implied volatility.

The following four hypotheses will be examined:

Hypothesis 1:
H_0: Implied volatility is not a significant predictor of future realised volatility, $\beta_1 = 0$
H_A: Implied volatility is a significant predictor of future realised volatility, $\beta_1 \neq 0$

Hypothesis 2:
H_0: Historical volatility is not a significant predictor of future realised volatility, $\beta_1 = 0$
H_A: Historical volatility is a significant predictor of future realised volatility, $\beta_1 \neq 0$

Hypothesis 3:
H_0: Implied volatility is not a significant predictor of future realised index returns, $\beta_1 = 0$
H_A: Implied volatility is a significant predictor of future realised index returns, $\beta_1 \neq 0$

Hypothesis 4:
H_0: The changes in implied volatility is not a significant predictor of future realised index returns, $\beta_1 = 0$
H_A: The changes in implied volatility is a significant predictor of future realised index returns, $\beta_1 \neq 0$

3.2 Scientific Perspective

The scientific perspective of a thesis can either be quantitative or qualitative where a quantitative method emphasises the quantification of data collection and testing theories whereas a qualitative method emphasises the generation of theories. These research methods can then be divided into various approaches called ontology, epistemology and finally methodology. Ontology is the view of the reality where knowledge is seen to be either an objective reality (objectivism) or constructed by the owner (constructivism). Epistemology is the view of knowledge where the aim of the research is to either understand the laws that govern behaviour (positivism) or gain in-depth understanding of the research subject to understand why they do what they do (interpretivism). Lastly, methodology is the way of acquiring the knowledge and can be done either by a deductive or inductive approach (Bryman & Bell, 2011).

This thesis is based upon a quantitative research method as our emphasis lies on testing theories within the field rather than attempting to establish new theory. It also takes on a positivistic and objective approach as the aim is to attempt to understand the future behaviour of the stock market by analysing implied volatility. The hypotheses are deduced with the basis of known theory about implied volatility and then tested through statistical tests and regressions using real-world observations. This altogether shows the appropriateness of using a quantitative and deductive approach. As the qualitative research method emphasises the theoretical outcome of the research through an inductive approach as well as attempts to explain the behaviour of the studied objects it is thus an inappropriate method for this thesis.

3.3 Method

This study aims to answer two main research questions: if and how well implied volatility predicts realised returns as well as if and how well implied volatility predicts realised index returns. To answer these research questions and test our hypotheses six different regressions, four univariate and two multiple regressions, are performed through the Ordinary Least Squares (OLS) method. Each regression is performed using the statistical package Eviews.

For our data the natural logarithm is taken for each variable value since this makes the effective relationship non-linear, while still preserving the linear model. Taking the natural logarithm is also a means of transforming a highly skewed variable into one that is more approximately normal (Christensen & Prabhala, 1998). Yin and Moffatt (2019) also states the necessity of restricting the volatilities to be positive by taking the natural logarithm of the values since there is a possibility that the OLS-regression would provide negative fitted values. As volatilities cannot take negative values, these then negative fitted values would result in a meaningless prediction of option prices through the BSM-model. This is the case mainly because this would imply that the option premium is more negative than the cost of carry of the rates used to price the option. Hence, if this is the case it would mean that the implied volatility of the BSM-model would be impossible to define. Negative volatility is additionally counterintuitive as it would mean that option sellers are willing to sell the options with more negative expected return than the market interest rate.

The study is limited to the OMXS30 index and OMXS30 index options on the Nasdaq Stockholm Stock Exchange and only short-term relations of one month are studied. The data sample consists of monthly implied volatilities of OMXS30 index options and OMXS30 index returns between the 10th of May 2012 and 9th of February 2020, resulting in 92 monthly periods. We chose this monthly period since the implied volatility data that we could access from Thomas Reuters Datastream was available from the 10th of May 2012 and hence we resonated that the 10th is a logical starting point for every monthly period. Also, using a monthly period which starts the 10th and ends the 9th each month makes it possible to include every day of each month without having to make any adjustments for if each month has 28 to 31 days, like one would if the monthly period started at the first day of the month and ended the last. The returns are calculated from OMXS30 index closing prices downloaded from financial software Thomson Reuters Datastream. The OMXS30 closing prices and returns are used to calculate the past historical volatility and realised volatility. The implied volatility values are downloaded from Thomson Reuters Datastream.

3.3.1 Data Collection

The data set used in this study is obtained using the financial database Thomas Reuters Datastream. From this database we downloaded our data for our time period May 10th 2012 to February 9th 2020. The following data was acquired for this period: implied volatility data from continuous OMXS30 index call and put options with one month to maturity and OMXS30 index closing prices. OMXS30 is the leading share index of Stockholm Stock Exchange and consists of the 30 most traded stocks on the exchange (Nasdaq OMX, 2020). This index is traded weekly with European-style options, where options of OMXS30 expire each Friday except for the third Friday of each month (Nasdaq, no date).

To compute the realised volatility daily closing prices of the OMXS30 index was acquired from Thomson Reuters Datastream. An ex-post monthly realised volatility was computed by calculating the standard deviation of the daily closing prices within one months time, from the 10th each month to the 9th the following month. This volatility was then annualised to match the implied volatility (Equation 5). The historical volatility is the realised volatility one monthly period before. Hence, realised volatility lagged one month becomes historical volatility. The data collected for realised volatility and historical volatility are to be used together in an OLS-regression. An illustration of our data sampling procedure for historical volatility and realised volatility is presented below in figure 1. Historical volatility recorded for "Monthly period 1" and regressed with realised volatility that is recorded for "Monthly period 2". For the next monthly period the realised volatility for "Monthly period 2" becomes the historical volatility that is regressed with realised volatility for monthly period three and so on for 92 monthly periods.

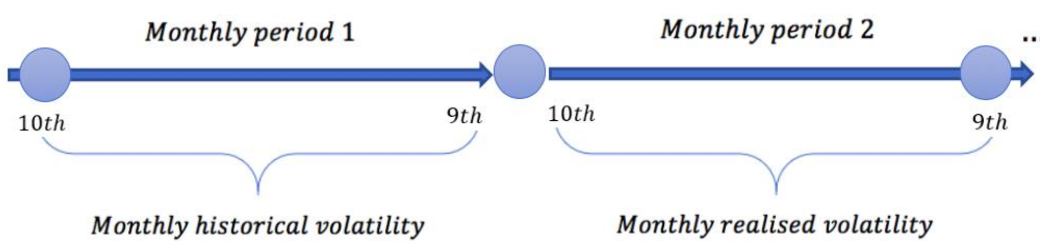

Figure 1. Illustrates how historical volatility and realised volatility data are collected for each monthly period.

The implied volatility used in this study is based on implied volatility data collected from Thomson Reuters Datastream. Implied volatility was downloaded monthly from continuous OMXS30 call and put options. The implied volatility data retrieved from Thomson Reuters Datastream is calculated using equation (8) and (9) from the Black-Scholes-Merton model. This calculation of implied volatility takes into account dividends paid out on the underlying asset. The continuous call and put options used to extract implied volatility from were at-the-money with one-month constant maturity.

Call and put options have separate implied volatility, so to calculate the implied volatility used for this study an unweighted average of the call and put option implied volatility was computed for each month. We decided to take the average implied volatility of call and put options considering that put options are traded almost as frequently as call options (Christensen & Hansen, 2002) and that implied volatility of call and put options are almost identical (Szakmary et al (2003). During our sample period from May 10th 2012 to April 9th 2020 the average trading volume per day was 28 call options and 26 put options (Thomson Reuters Datastream). Which we concluded to be equal enough to take a simple unweighted average of the implied volatility of the call and put options. Additionally, it would be a time-consuming process to take the trade weighted average throughout our sample period. We decided to use continuous option series since they provide an uninterrupted view of implied volatility over time. The continuous options series available on Thomson Reuters Datastream does not, unlike regular individual options series, expire until the actual options class stops existing. We decided to only use options with one month left to maturity in order to get implied volatilities that provide the market's views about the following month. Finally, we chose to use at-the-money options since options that are too far from being at the money may provide less useful information about the implied volatility as at-the-money options are more actively traded (Jiang & Tian, 2005). Additionally, Szakmary et al (2003) and Canina and Figelwski (1993) found implied volatility from at-the-money options to be better predictors of future realised volatility compared to implied volatility of in- or out-of-the-money options.

The implied volatility, as an unweighted average of call and put implied volatility, was lagged by one month to act as a predictor for realised volatility. Implied volatility is recorded on the first day of the monthly period, on the 10th, and realised volatility is estimated at the end of the month for the monthly period ranging from the 10th each month to the 9th the following month. By doing this, each monthly period is represented by one ex-ante implied volatility forecast and one ex-post realised volatility computation. Hence, there is a series of non-overlapping data to avoid the overlapping issues, discussed in Christensen and Prabhala (1998), that could cause strong autocorrelation and possible regression errors. The data collected for realised volatility and implied volatility are to be used together in an OLS-regression. An illustration of our data sampling procedure for implied volatility and realised volatility is presented below in figure 2. Monthly implied volatility is calculated for "Monthly period 1" which is regressed against the monthly realised volatility for "Monthly period 2", then monthly implied volatility is calculated for "Monthly period 2" which is regressed against the next monthly period and so on for a total of 92 monthly periods.

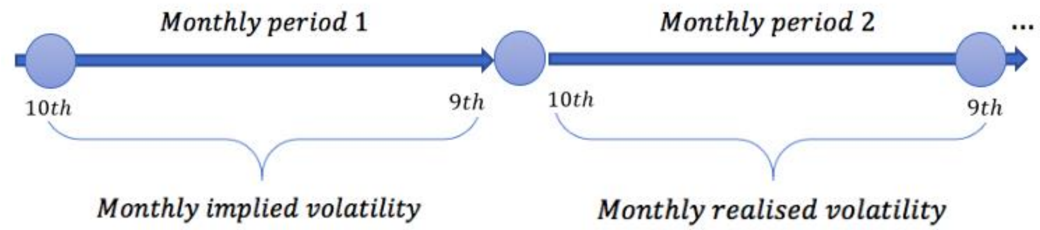

Figure 2. Illustrates how implied volatility and realised volatility data are collected for each monthly period.

The daily realised returns are calculated using daily OMXS30 index closing prices by applying equation (7) and the change in implied volatility is computed as the relative daily change in the implied volatility data by applying the following equation:

$$\Delta IV_t = \ln\left(\frac{IV_t}{IV_{t-1}}\right) \tag{10}$$

Where ΔIV_t, IV_t and IV_{t-1} is the change in implied volatility, implied volatility at time t and implied volatility at time t-1 respectively. Monthly averages of returns and the changes in implied volatility are computed from 10th to 9th the following month. In order for implied volatility and the change in implied volatility to act as predictors of realised returns in an OLS-regression they are both lagged one month against the realised returns. Illustrations of our data sampling procedure for implied volatility and realised returns as well as for change in implied volatility and realised returns are presented below, in figure 3 and figure 4 respectively. Monthly implied volatility is calculated for "Monthly period 1" which is regressed against the monthly average realised returns for "Monthly period 2", then monthly implied volatility is calculated for "Monthly period 2" and is regressed with monthly average realised returns for the next monthly period and so on for 92 monthly periods.

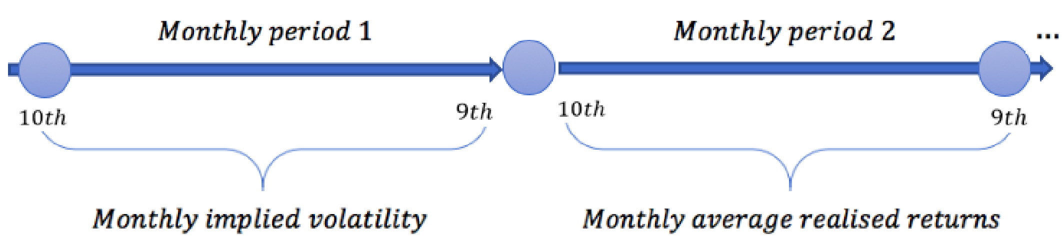

Figure 3. Illustrates how implied volatility and realised returns data are collected for each monthly period.

Monthly change in implied volatility is calculated for "Monthly period 1" which is regressed against the monthly average realised returns for "Monthly period 2", later monthly change in implied volatility is calculated for "Monthly period 2" and is regressed with monthly average realised returns for the next monthly period and so on for 92 monthly periods.

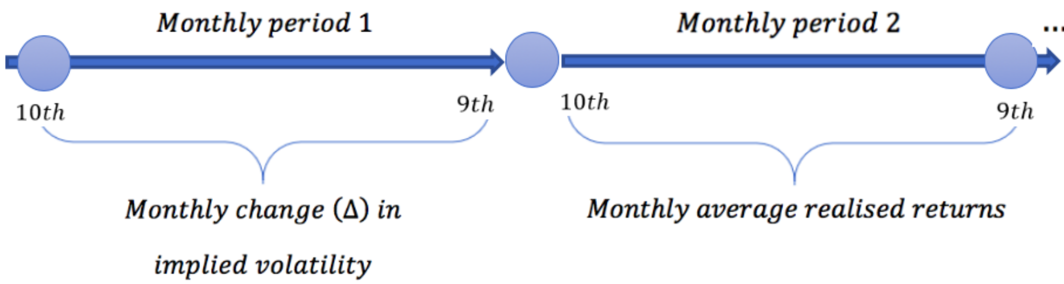

Figure 4. Illustrates how changes in implied volatility and realised returns data are collected for each monthly period.

3.3.2 Ordinary Least Squares Assumptions

In order to use the Ordinary Least Squares (OLS) regression model it is required that the model tested is linear, that the relationship between the independent variable and dependent variable can be expressed diagrammatically by a straight line and that the parameters are not divided, multiplied together, squared and so on (Brooks, 2008). Apart from this requirement, there are further multiple assumptions for OLS-regressions which are the following (ibid.):

1. The error terms have a zero mean
2. The variance of the error terms is constant and finite over all values of x
3. The error terms are linearly independent of one another
4. There is no relationship between the error term and its corresponding x-variable
5. The error term is normally distributed

Assuming that the above-mentioned assumptions hold, the OLS-estimator is shown to have properties of being consistent, efficient and unbiased (Brooks, 2008). Unbiased refers to that the estimated values for the coefficients will be equal to their true values and efficient refers to minimising the probability that the estimates are off from their true values.

However, if these assumptions do not hold, one will encounter several problems when performing the regression analysis which may affect the reliability or accuracy of the results. The main problems that will be addressed in this thesis are multicollinearity, heteroscedasticity and autocorrelation which all implies a violation of one of these assumptions.

Multicollinearity refers to the assumption of a linear model and infers a strong correlative relationship between the explanatory variables. When there is multicollinearity between the independent variables the estimation for all coefficients may be wrong as the variables contain information to estimate only one parameter (Brooks, 2008). To test for multicollinearity, one can either construct a correlation matrix or perform a Variance Inflation Factors (VIF) test. In this thesis, a VIF-test will be performed to determine the strength of the correlative relationship between the variables in our multiple regression models.

The other issue, heteroscedasticity, refers to the second assumption of errors having constant variance (homoscedasticity), which means that the error terms do not have constant variance. Whilst this does not affect the OLS-estimates and their attributes of being unbiased and consistent it does however cause them to be inefficient and any inferences made upon these estimates may be misleading (Brooks, 2008). There are various tests to see whether the errors are homoscedastic or heteroscedastic and a common one is the Breusch-Pagan test as well as the White's test. In this thesis, a White's test will be performed as this test is considered to be more robust than the Breusch-Pagan test (ibid.).

Lastly, autocorrelation refers to the third assumption implying that the error terms are uncorrelated with one another and can be either positive or negative (Brooks, 2008). One way to test for autocorrelation is to plot the residuals and then examine those graphs. However, it may be difficult to interpret so we therefore find a formal statistical test to be more appropriate and will thus perform a Durbin-Watson test. The Durbin-Watson test examines the relationship between an error term at time t and its immediate value before that, the error term at time $t-1$ (ibid.).

Furthermore, another issue worth discussing is the violation of the fifth assumption, the normality assumption. The inferences made based on the OLS-estimates may be wrong if this is violated (Brooks, 2008). If the errors are not normally distributed it would indicate that the explanatory variables in the regression does not have the same explanatory ability throughout the whole data set of the dependent variable (ibid.). This can be tested for using a Jarque-Bera test which will be conducted in this thesis.

We do not find it to be necessary to test the first assumption as it will never be violated if a constant (the intercept) is included in the regression model (Brooks, 2008). We do however find it to be necessary to test for multicollinearity, heteroscedasticity, autocorrelation and the normality assumption to ensure the applicability of our data.

3.3.3 Empirical Models

Regression model I is used to study hypothesis 1, if and how well implied volatility predicts future realised volatility:

Model I:

$$lnRV_t = \alpha_0 + \beta_1 lnIV_{t-1} + \varepsilon_t$$

Regression model II is used to study hypothesis 2, if and how well historical volatility predicts future realised volatility:

Model II:

$$lnRV_t = \alpha_0 + \beta_1 lnRV_{t-1} + \varepsilon_t$$

Regression model III is used to study if and how well implied volatility together with historical volatility predicts future realised volatility:

Model III:

$$lnRV_t = \alpha_0 + \beta_1 lnIV_{t-1} + \beta_2 lnRV_{t-1} + \varepsilon_t$$

Where α_0, $\beta_1$and β_2, IV_{t-1}, RV_t, $RV_{t-1}$and ε_t is the intercept, slope coefficients, lagged implied volatility, realised volatility, historical volatility and the error term respectively at time t.

Regression model IV is used to study hypothesis 3, if and how well implied volatility predicts realised returns:

Model IV:

$$r_t = \alpha_0 + \beta_1 \ln IV_{t-1} + \varepsilon_t$$

Regression model V is used to study hypothesis 4, if and how well the change in implied volatility predicts realised returns:

Model V:

$$r_t = \alpha_0 + \beta_1 \Delta IV_{t-1} + \varepsilon_t$$

Regression model VI is used to study if and how well implied volatility and the change in implied volatility predicts realised returns:

Model VI:

$$r_t = \alpha_0 + \beta_1 \ln IV_{t-1} + \beta_2 \Delta IV_{t-1} + \varepsilon_t$$

Where r_t, α_0, $\beta_1$and β_2,IV_{t-1}, ΔIV_{t-1} and ε_t is the return at time t, intercept, slope coefficients, lagged implied volatility, lagged change in implied volatility and the error term at time t respectively.

3.4 Reliability and Validity

When examining a study's quality reliability, replicability and validity are three criterions that are necessary to address and are particularly important to discuss when conducting a quantitative study (Bryman & Bell, 2011). As these criteria are associated with the measures used in quantitative studies that may be inconsistent or may be seen as unreliable it is of high importance to discuss our study out of these viewpoints.

The concept of reliability refers to the replicability of a study's results where a study is considered to be replicable and will reach the same conclusion as long as the chosen sample and time frame are the same (Bryman & Bell, 2011). Furthermore, reliability depends upon the measure's stability over time stating that the same results would be achieved regardless of the amount of times it was used. As our data is secondary publicly available data consisting of historical index prices as well as implied volatilities, we believe that the replicative ability of our thesis is strong. Hence, we believe that one will reach the same conclusion as long as the same presented time period and sample is used. Furthermore, to improve the reliability of a study one can perform winsorization or trimming which are methods based upon the fifth OLS-assumption. The methods are means of either replacing the extreme values to be closer to the mean value (winsorization) or removing the extreme values (trimming). However, it is further stated according to the Central Limit Theorem that if the sample contains over 30 observations the OLS assumption that error terms should be normally distributed is not of essence compared to when the sample size is smaller, as the extreme values are said to have less impact when the sample size is large (Brooks, 2008). In conjunction with this statement we did therefore not find it to be necessary to modify our data through winsorization or trimming since the observations which would be excluded through these processes could contain valuable information. We believe that excluding these data points could bias our results and thus decided to run our regressions on the original data as our assessment is that using original data would make our thesis more reliable compared to using adjusted data.

Validity refers to whether a measure of a concept really measures that concept and is considered to be the most important criterion as it presumes reliability (Bryman & Bell, 2011). As our chosen method is based on previous research using measures equal to the ones that we have presented in the preceding chapters we believe that they are reliable and measure what they are intended to measure.

3.5 Source Critical Consideration

The data which this thesis is based upon is acquired from Thomson Reuters Datastream and is thus secondary data. Although secondary data may not be as reliable as primary data, since users lack knowledge of how the data in detail have been collected, we believe that this data is of high quality and trustworthy. Furthermore, as Datastream is a database extensively used in both industry and academia and Thomson Reuter being a well-known organisation it is motivated that the data is reliable as well as accurate.

The theoretical framework consists of both academic peer-reviewed journals and published literature. Factual information presented in the literature review have been collected from published literature written by researchers and is used in academia as educational literature. As the literature which we reference upon does not consist of chapters written by other authors nor is in turn referencing other sources we thus believe that these are reliable sources. Empirical results within the field have entirely and exclusively been collected from academic peer-reviewed journals. As academic peer-reviewed journals are written by experienced researchers within the field and then reviewed by multiple other experienced researchers to ensure the article's quality they can be seen as more reliable than other sources, such as other theses. We therefore believe that the sources used in this thesis are trustworthy, accurate and unbiased.

3.6 Research Ethical Reflection

When conducting a study, it is of high importance as a researcher to discuss the possible ethical complications one may encounter and create. Some of these ethical issues are associated with the data collection while others are associated with the research subjects such as the risk of deception, causing harm, invasion of privacy and so on. Furthermore, in line with the ethical perspective it is of high importance to avoid personal interests and to present objective as well as honest results.

This thesis is, as mentioned, conducted in a quantitative manner and any ethical issues concerning research subjects is, based on this, evidently not expected. Considering that our data is collected from Thomson Reuters Datastream which contains publicly available data we do not expect any ethical issues concerning this data. Neither does our data reveal any confidential information and the results in this thesis are presented in an objective manner in the sense that the results will not be manipulated to favour a certain outcome.

4. Analysis and Findings

4.1 Descriptive Statistics

Table 4.1 provides a description of key statistics of the variables used in our regression models. This summary includes the minimum respective maximum values of the variables, the standard deviation, the mean as well as the median of the data set. The table also contains a description of kurtosis as well as skewness which measures the number of outliers present in the distribution and the extent of which a variable's distribution is not symmetrical about the mean respectively.

Table 4.1

	Realised Vol	*Implied volatility*	*Historical volatility*	*Ln Realised Volatility*
Mean	0.146604526	0.170784678	0.148081791	-1.991887318
Median	0.130266103	0.160538097	0.130266103	-2.038185228
Standard Deviation	0.059974914	0.041308648	0.061751452	0.381434815
Kurtosis	4.892365077	0.582091586	4.157990731	3.052326213
Skewness	1.785535196	1.093966262	1.696421832	-0.230689764
Minimum	0.027955708	0.114552613	0.027955708	-3.57713388
Maximum	0.419161199	0.300097578	0.419161199	-0.869499711
Observations	92	92	92	92

	Ln Implied volatility	*Ln Historical Volatility*	*Returns*	*Delta Implied volatility*
Mean	-1.793439922	-1.984937039	0.00030791	-0.000347191
Median	-1.829231379	-2.038185228	0.000415269	-0.00050125
Standard Deviation	0.224429431	0.38914636	0.00180026	0.014707228
Kurtosis	-0.315837179	2.749703439	0.503149583	5.800288799
Skewness	0.674032249	-0.193105976	-0.517509138	-0.90046795
Minimum	-2.166721062	-3.57713388	-0.004956157	-0.071963389
Maximum	-1.203647598	-0.869499711	0.003972821	0.041481373
Observations	92	92	92	92

A normally distributed variable should theoretically have a kurtosis at 3 and a skewness at 0. The descriptive statistics reveals that the only variable that is somewhat normally distributed is the natural logarithm of realised volatility. Judging by the table we can observe that the variables before taking the natural logarithm all possess characteristics of high skewness as well as high kurtosis, except from implied volatility which only has high skewness. This means that the data of realised and historical volatility, before taking the natural logarithm, were highly skewed with a lot of outliers and that implied volatility had a lack of outliers, although it being highly skewed as well. Furthermore, although financial series typically in practice are characterised by a leptokurtic (kurtosis greater than 3) distribution (Brooks, 2008), we discover mixed results. We can see that the variables used in our regressions partially are characterised by a platykurtic (kurtosis less than 3) distribution and partially by a leptokurtic distribution. We can also see that the change in implied volatility is highly leptokurtic which indicates a large number of outliers.

4.2 Tests for Multicollinearity, Heteroscedasticity, Autocorrelation and Normality

To interpret the VIF test, which can be found in the Appendix, the column “Centered VIF” is of interest as this determines the level of correlation which we can interpret using the following conditions:

VIF	Degree of correlation
> 5	Highly correlated
1-5	Moderately correlated
1	Not correlated

These conditions imply that a VIF > 5 warrants further investigation but is however not as severe as a VIF > 10. A VIF level of >10 requires adjustments since it indicates severe problems with multicollinearity. The results from our conducted VIF tests on our two multiple regressions reveal that our explanatory variables are not correlated to moderately correlated in models III and VI, with a VIF of 1.4 for model III and 1.0 for model VI. We can therefore conclude that there is no apparent need to adjust for multicollinearity in our data.

The null hypothesis for the White’s test is that the errors are homoscedastic and hence, if we reach the conclusion to reject the null the errors are thus concluded to be heteroscedastic. The test produces three different test statistics, “F-statistic”, “Obs*R-Squared” and “Scaled Explained SS” as well as the probability of each test statistic. The results reveal that we fail to reject the null hypothesis at the 5% significance level for all models and hence we conclude that the errors are homoscedastic.

To test whether our data contains autocorrelation or not the Durbin-Watson test statistic will be examined to see whether we reject or fail to reject the null hypothesis stating that there is no evidence of autocorrelation. This test statistic will always be a value between 0 and 4 and below in figure 5 are the following decision rules as well as regions used to assess whether we can reject or fail to reject the null hypothesis:

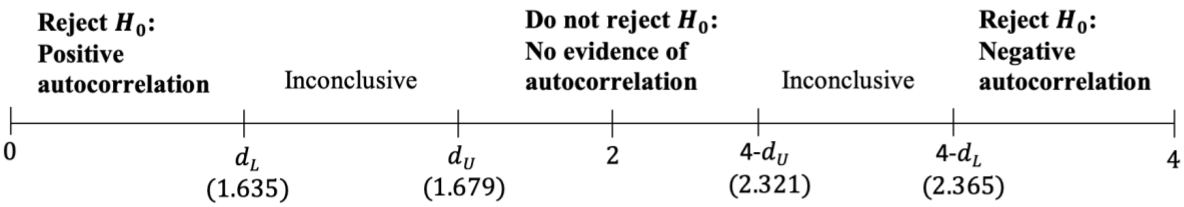

Figure 5.

Where d, d_L and d_U is the Durbin-Watson test statistic, the lower bound critical value and the upper bound critical value respectively. From this, we hence fail to reject the null hypothesis and thus conclude that there is no autocorrelation in the regression models II, III, IV, V and VI. Model I however, falls within the inconclusive region and is thus making it impossible to conclude whether it contains autocorrelation or not. As it falls within the inconclusive region, we will modify the test for this specific model as described by Harrison (1975) and reject the null hypothesis as d is less than d_U and conclude that we have autocorrelation. We will furthermore conclude that autocorrelation for model I is positive, as the Durbin-Watson statistic for this model is between 0-2.

To test for the normality assumption, Jarque-Bera tests have been conducted with the null hypothesis stating that the errors are normally distributed. The test reveals that the probability for the Jarque-Bera test statistic is less than the significance level of 5% for models I, II and III and greater than the significance level of 5% for models IV, V and VI. Hence, we reject the null hypothesis for models I, II and III whereas we fail to reject the null hypothesis for IV, V and VI and thus conclude that the errors in the models I-III are not normally distributed while the errors in models IV-VI are. To obtain normally distributed errors in models I-III a 95% winsorization of the data could be performed. However, as previously discussed in section 3.4, we decided to not winsorize our data to not lose valuable information in our data set.

4.3 Results

4.3.1 Regression Analysis of Implied Volatility and Realised Volatility

The OLS regression results of regression model I can be viewed below in table 4.2. For this regression we try to test hypothesis 1, if implied volatility is a significant predictor of future realised volatility. The implied volatility variable has a slope coefficient $\beta_1$of 0.507 and the respective p-value is 0.004. This means that the slope coefficient $\beta_1$is significantly different from zero and the null hypothesis of hypothesis 1, that β_1= 0, can be rejected at a 1% significance level. Hence, the conclusion that implied volatility is a significant predictor of realised volatility can be made. The adjusted R^2 of the regression is 0.079 which means that 7.9% of the variation of the dependent variable, realised volatility, is explained by the independent variable, implied volatility.

Table 4.2

Variable	Coefficient	Std. Error	t-Statistic	p-value	Model Summary	
Intercept	-1.083	0.309	-3.504	0.001	R-squared	0.089
					Adj. R-squared	0.079
Implied Volatility	0.507	0.171	2.964	0.004***	S.E. of regression	0.366
					Durbin-Watson	1.646

Table 4.2 reports the OLS estimates for the univariate regression model I where implied volatility is the independent variable and future realised volatility is the dependent variable. The slope coefficient β_1, test statistic, test statistic and p-value (significance level) is reported for the independent variable. The summary statistics includes values of R-squared, adjusted R^2, S.E. of regression (standard error of regression) and Durbin-Watson test statistic for the regression. Different significance levels are indicated by *, ** and *** where they indicate a 10%, 5% and 1% significance level respectively. Regression model I: $lnRV_t = \alpha_0 + \beta_1 lnIV_{t-1} + \varepsilon_t$

4.3.2 Regression Analysis of Historical Volatility and Realised Volatility

The regression results of regression model II can be viewed below in table 4.3. For this regression we try to test hypothesis 2, if historical volatility is a significant predictor of future realised volatility. The historical volatility variable has a slope coefficient β_1 of 0.317 and the respective p-value is 0.002. This indicates that the slope coefficient β_1 is significantly different from zero and the null hypothesis of hypothesis 2, that $\beta_1 = 0$, can be rejected at a 1% significance level. Hence, the conclusion that historical volatility is a significant predictor of realised volatility is made. The adjusted R^2 of the regression is 0.095 which means that 9.5% of the variation of the dependent variable, realised volatility, can be explained by the independent variable, historical volatility. Comparing these results to table 4.2 we can draw the conclusion that historical volatility has a higher R^2 and that historical volatility hence has higher explanatory power of future realised volatility compared to implied volatility.

Table 4.3

Variable	Coefficient	Std. Error	t-Statistic	p-value	Model Summary	
Intercept	-1.362	0.198	-6.891	0.000	R-squared	0.105
					Adj. R-squared	0.095
Historical Volatility	0.317	0.098	3.244	0.002***	S.E. of regression	0.363
					Durbin-Watson	2.215

Table 4.3 reports the OLS estimates for the univariate regression model II where historical volatility is the independent variable and future realised volatility is the dependent variable. The slope coefficient β_1, standard error, test statistic and p-value (significance level) is reported for the independent variable. The summary statistics includes values of R^2, adjusted R^2, S.E. of regression (standard error of regression) and Durbin-Watson test statistic for the regression. Different significance levels are indicated by *, ** and *** where they indicate a 10%, 5% and 1% significance level respectively. Regression model II: $lnRV_t = \alpha_0 + \beta_1 lnRV_{t-1} + \varepsilon_t$

4.3.3 Regression Analysis of Implied Volatility, Historical Volatility and Realised Volatility

In table 4.4 we try to find out how well implied volatility and historical volatility predict realised volatility together in a multiple regression. The regression results of model III are hence displayed in table 4.4. From the multiple regression results we can see that the implied volatility slope coefficient β_1 is 0.302 and the respective p-value is 0.130. This means that the implied volatility slope coefficient $\beta_1$is not significantly different from zero at a 1%, 5% or 10% significance level in this multiple regression. Contrarily, historical volatility is significant at a 5% significance level with a slope coefficient $\beta_2$of 0.226 and a p-value of 0.05. Hence, implied volatility is not a significant predictor in this multiple regression, while historical volatility is. The adjusted R^2 for this multiple regression is 0.128, meaning that the independent variables, implied volatility and historical volatility, together explain 12.8% the variation of the dependent variable, realised volatility. This adjusted R^2 value is greater than the univariate regressions in table 4.2, with implied volatility as the independent variable, and table 4.3, with historical volatility as the independent variable. Hence, a multiple regression with both implied volatility and historical volatility as explanatory variables outperform both univariate regressions where these variables are used separately.

Table 4.4

Variable	Coefficient	Std. Error	t-Statistic	p-value	Model Summary	
Intercept	-1.002	0.307	-3.267	0.002	R-squared	0.128
					Adj. R-squared	0.108
Implied Volatility	0.302	0.198	1.527	0.130	S.E. of regression	0.360
					Durbin-Watson	2.127
Historical Volatility	0.226	0.114	1.985	0.05**		

Table 4.4 reports the OLS estimates for the multiple regression model III where implied volatility and historical volatility are the independent variables and future realised volatility is the dependent variable. The slope coefficients β_1 and β_2, standard error, test statistic and p-value (significance level) is reported for the independent variables. The summary statistics includes values of R-squared, adjusted R^2, S.E. of regression (standard error of regression) and Durbin-Watson test statistic for the regression. Different significance levels are indicated by *, ** and *** where they indicate a 10%, 5% and 1% significance level respectively. Regression model III: $lnRV_t = \alpha_0 + \beta_1 lnIV_{t-1} + \beta_2 lnRV_{t-1} + \varepsilon_t$

4.3.4 Regression Analysis of Implied Volatility and Realised Returns

The regression results of regression model IV can be viewed below in table 4.5. This regression investigates hypothesis 3, if implied volatility is a significant predictor of future realised returns. The implied volatility has a slope coefficient β_1 of 0.001 with a respective p-value of 0.470. This p-value indicates that the slope coefficient β_1 is not significantly different from zero and that we hence fail to reject the null hypothesis, that $\beta_1 = 0$, at a 1%, 5% as well as a 10% significance level. We thus conclude that implied volatility is not a significant predictor of future realised returns. The adjusted R^2 for this regression is negative at -0.005 or -0.5%, meaning that implied volatility has almost no explanatory power of the movements in realised returns and further implies that implied volatility is a poor predictor of returns. A low or negative R^2 value means that the model being used is not a good fit for the data (Davidson & Levin, 2014).

Table 4.5

Variable	Coefficient	Std. Error	t-Statistic	p-value	Model Summary	
Intercept	0.001	0.002	0.922	0.359	R-squared	0.006
					Adj. R-squared	-0.005
Implied Volatility	0.001	0.001	0.725	0.470	S.E. of regression	0.002
					Durbin-Watson	2.284

Table 4.5 reports the OLS estimates for the univariate regression model IV where implied volatility is the independent variable and future realised return is the dependent variable. The slope coefficient β_1, test statistic, standard error and p-value (significance level) is reported for the independent variable. The summary statistics includes values of R^2, adjusted R^2, S.E. of regression (standard error of regression) and Durbin-Watson test statistic for the regression. Different significance levels are indicated by *, ** and *** where they indicate a 10%, 5% and 1% significance level respectively. Regression model IV: $r_t = \alpha_0 + \beta_1 lnIV_{t-1} + \varepsilon_t$

4.3.5 Regression Analysis of Changes in Implied Volatility and Realised Returns

Regression model V results can be viewed below in table 4.6. This regression investigates hypothesis 4, if changes in implied volatility is a significant predictor of future realised returns. The delta implied volatility has a slope coefficient β_1 of 0.017 with a respective p-value of 0.188. This p-value indicates that the slope coefficient β_1 is not significantly different from zero and that we hence fail to reject the null hypothesis, that $\beta_1 = 0$, at a 1%, 5% and 10% significance level. We thus conclude that delta implied volatility is not a significant predictor of future realised returns. The adjusted R^2 for this regression is 0.008, meaning that changes in implied volatility explain 0.8% of the variation of the dependent variable, realised returns. Comparing these regression results to the regression results in table 4.5, with implied volatility as an explanatory variable for realised returns, we can conclude that although delta implied volatility is not significant at a 10% significance level it is a more significant estimate than implied volatility. Additionally, delta implied volatility has a higher adjusted R^2 value at 0.8%, compared to implied volatility with an adjusted R^2 value of -0.5%, which means that delta implied volatility is slightly superior to implied volatility in predicting realised returns.

Table 4.6

Variable	Coefficient	Std. Error	t-Statistic	p-value	Model Summary	
Intercept	0.0003	0.0002	1.678	0.097	R-squared	0.019
					Adj. R-squared	0.008
Delta Implied Volatility	0.017	0.013	1.327	0.188	S.E. of regression	0.002
					Durbin-Watson	2.199

Table 4.6 reports the OLS estimates for the univariate regression model V where delta implied volatility is the independent variable and future realised return is the dependent variable. The slope coefficient β_1, test statistic, standard error and p-value (significance level) is reported for the independent variable. The summary statistics includes values of R^2, adjusted R^2, S.E. of regression (standard error of regression) and Durbin-Watson test statistic for the regression. Different significance levels are indicated by *, ** and *** where they indicate a 10%, 5% and 1% significance level respectively. Regression model V: $r_t = \alpha_0 + \beta_1 \Delta IV_{t-1} + \varepsilon_t$

4.3.6 Regression Analysis of Implied Volatility, Changes in Implied Volatility and Realised Returns

In table 4.7 we try to find out how well implied volatility and changes in implied volatility predicts realised returns together in a multiple regression. The regression results of model VI are hence displayed in table 4.7. From the multiple regression results we can see that the implied volatility slope coefficient $\beta_1$is 0.001 and the respective p-value is 0.583. This means that the implied volatility slope coefficient $\beta_1$is not significantly different from zero at a 1%, 5% or 10% significance level in this multiple regression. Delta implied volatility has a slope coefficient $\beta_2$of 0.016 respectively a p-value of 0.221, indicating that delta implied volatility is not a significant predictor either in this multiple regression at a 1%, 5% or 10% significance level. However, delta implied volatility is more significant than implied volatility. The adjusted R^2 for this multiple regression is 0.001, meaning that the independent variables, implied volatility and delta implied volatility, together explain 0.1% the variation of the dependent variable, realised returns. Hence, this multiple regression performs worse in explaining realised returns compared to the univariate regression in table 4.6, where changes in implied volatility is the sole independent variable. Additionally, this multiple regression outperforms the regression in table 4.5, where implied volatility is used as the sole independent variable, in explaining realised returns. In conclusion, no independent variable in this multiple regression was significant at a 1%, 5% or 10% significance level. We can therefore conclude that the univariate regression with delta implied volatility in table 4.6 outperforms this multiple regression in explaining realised returns.

Table 4.7

Variable	Coefficient	Std. Error	t-Statistic	p-value	Model Summary	
Intercept	0.001	0.002	0.752	0.454	R-squared	0.023
					Adj. R-squared	0.001
Implied Volatility	0.001	0.001	0.552	0.583	S.E. of regression	0.002
					Durbin-Watson	2.186
Delta Implied Volatility	0.016	0.013	1.234	0.221		

Table 4.6 reports the OLS estimates for the multiple regression model VI where implied volatility and delta implied volatility are the independent variables and future realised return is the dependent variable. The slope coefficient β_1 and β_2, test statistic, standard error and p-value (significance level) is reported for the independent variables. The summary statistics includes values of R^2, adjusted R^2, S.E. of regression (standard error of regression) and Durbin-Watson test statistic for the regression. Different significance levels are indicated by *, ** and *** where they indicate a 10%, 5% and 1% significance level respectively. Regression model VI:$r_t = \alpha_0 + \beta_1 lnIV_{t-1} + \beta_2 \Delta IV_{t-1} + \varepsilon_t$

5. Discussion and Critical Reflection

This paper aims to answer our two main research questions about implied volatility: if and how well it can predict 1) realised volatility and 2) realised returns. This section will discuss the empirical findings of the six different regressions models based on our research question and hypotheses. These empirical findings from the regressions will be discussed and compared against each other as well as with the previous research presented in the literature survey section. Findings in previous research regarding implied volatility as a predictor have been of mixed results. The most frequent conclusions of the papers studied were that: 1) implied volatility is a suitable predictor of realised volatility that is superior to historical volatility in that matter and 2) implied volatility is not a suitable predictor of future returns since implied volatility and returns have a strong negative asymmetric relationship.

Chistensen and Prahbala (1998) examined how well monthly implied volatility of at-the-money call options on the S&P 100 index could predict realised volatility. Their study concluded that implied volatility is a significant predictor for realised volatility and is a superior predictor to historical volatility. Chistensen and Prahbala (1998) got the following adjusted R^2 values on regressions with realised volatility as the dependent variable: 39% for a regression with implied volatility as the single independent variable, 32% for a regressions with historical volatility as the single independent variable, 41% for a multiple regression with implied volatility and historical volatility as independent variables. Hence, a multiple regression with implied volatility and historical volatility had the highest adjusted R^2 value in their study.

Christensen and Hansen (2002) used non-overlapping observations, similar to Chistensen and Prahbala (1998), and studied the relationship between realised volatility and implied volatility of S&P 100 options. The implied volatility used was the trade weighted implied volatility of in-the-money and out-of-the-money call and put options. Their study concluded that implied volatility is a better and more significant predictor of realised volatility than historical volatility. Their empirical results from their regressions with realised volatility as the dependent variable showed an adjusted R^2 of: 25.08% for a regression with trade weighted implied volatility from call and put options as the independent variable and 25.01% for a multiple regression with this trade weighted implied volatility and historical volatility as the independent variables. Hence, this paper shows that implied volatility in a univariate regression shows higher adjusted R^2 values than a multiple regression with implied volatility and historical volatility together.

Szakmary et al (2003) studied data from futures options and tested how well implied volatility from at-the-money call and put options could predict future volatility of the underlying futures. The study used overlapping observations and corrections of standard errors of coefficients were made. The conclusion of the paper was that implied volatility was superior to historical volatility in predicting realised volatility for a large majority of the studied futures. The paper obtained the following adjusted R^2 values for their regressions on S&P 500 futures volatility and S&P 500 futures options with realised volatility as the dependent variable: 23.1% for a regression with implied volatility as the independent variable, 13.2% for a regression with historical volatility as the independent variable and 23.1% for a multiple regression on implied volatility and historical volatility.

The remainder of the papers studied concerning implied volatility as a predictor for realised volatility, namely Day and Lewis (1993) and Canina and Figelwski (1993), had lower values of R^2 for their regressions. Day and Lewis (1993) studied weekly implied volatility derived from near-the-money S&P 100 index call options as a predictor of realised volatility. These regressions made with realised volatility as the dependent variable showed and adjusted R^2 levels of: 3.7% for a regression with implied volatility as the independent variable and 9.4% for a regression with historical volatility as the independent variable. This paper concluded that implied volatility is a significant estimate that contains information about future volatility, but that it performs equally to forecasts from historical volatility.

Canina and Figelwski (1993) studied daily implied volatility of S&P 100 at-the-money call options. Their empirical results showed that implied volatility was significant only in 6 out of the 32 subsamples at a 5% significance level. The R^2 for these significant regressions of implied volatility as an independent variable ranged between 3.5-6.7%. The R^2 for regressions where realised volatility for the remaining life of options, with a maturity date of 7-35 days and for at-the-money and in-the-money options (with an intrinsic value of $0-$5), as the dependent variable is: 3.5% for a regression with implied volatility as the independent variable, 10.8% for a regression with historical volatility as the independent variable and 10.9% for a multiple regression with implied volatility and historical volatility as independent variables. According to Chistensen and Prahbala (1998), the results of Day and Lewis (1993) and Canina and Figelwski (1993) are flawed since they have used overlapping data which can lead to autocorrelation and regression errors.

The empirical results that was obtained from our study was that: implied volatility was statistically significant estimate at a 1% significance level and the adjusted R^2 of the univariate regression with implied volatility was 8.9% (Table 4.2), historical volatility was statistically significant estimate at a 1% significance level and the adjusted R^2 of the univariate regression with historical volatility was 10.5% (Table 4.3), in the multiple regression implied volatility was not a statistically significant estimate at a 10% significance level and historical volatility was statistically significant at a 5% significance level and the multiple regression adjusted R^2 was 12.8% (Table 4.4)

Comparing our regression results to previous studies we can observe that it is in accordance with the studies of Chistensen and Prahbala (1998), Christensen and Hansen (2002), Szakmary et al (2003) and Day and Lewis (1993) in that implied volatility is a statistically significant predictor of realised volatility. However, we find that our results are more coherent with Day and Lewis (1993) who conclude that implied volatility contains information about future volatility, but that it performs equally to forecasts from historical volatility. Our regressions adjusted R^2 is similar to the levels of Day and Lewis (1993) and Canina and Figelwski (1993) withR^2values around 4% for regressions with independent variable implied volatility, around 10% for historical volatility and Canina and Figelwski (1993) found an R^2 around 11% for a multiple regression on implied volatility and historical volatility. Compared to the adjusted R^2 of Chistensen and Prahbala (1998), Christensen and Hansen (2002) and Szakmary et al (2003) where the adjusted R^2 values are around: 23-39% for regressions with implied volatility as an independent variable, Chistensen and Prahbala (1998) and Szakmary et al (2003) found adjusted R^2 levels of around 13-32%, multiple regressions of implied volatility and historical volatility ranged from 23-41%. However, Szakmary et al (2003) adjusted R^2 of 13.2% for the regression on historical volatility is in line with our empirical results.

Dennis et al (2006) studied the relationship between implied volatility and returns and found evidence similar to Bekiros et al (2017), that there is a negative as well as asymmetric relationship between the two variables. Furthermore, Dennis et al's regression results show a $R^2$of 3.4% for all 50 firms in the sample as well as that the change in implied volatility is a poor predictor of returns. Amman, Verhofen and Süss (2008) examined whether implied volatility can predict future returns or not and discovered evidence of it being an appropriate predictor and found a positive relationship between the variables. The authors used returns as the dependent variable and lagged implied volatility as the independent variable and their regression results show an R^2 of 0.8%. Furthermore, in line with Amman, Verhofen and Süss's conclusion, Rubbaniy et al (2014) also discovered evidence of a positive relationship between implied volatility and future returns. They further report a very low R^2, although not stating the specific number, and that implied volatility is an appropriate predictor of future returns.

The empirical results that were discovered in our thesis are the following: implied volatility nor the change in implied volatility were statistically significant at a 1%, 5% or 10% significance level in univariate regressions and the adjusted R^2were -0.5% and 0.8% respectively. The variables were furthermore not statistically significant in a multiple regression at either level of significance and the adjusted $R^2$was 0.1%. These findings therefore support the conclusion that implied volatility as well as the change in implied volatility are not significant predictors of future returns which is in line with what was concluded by Dennis et al (2006) and Egbers and Swinkels (2015). Additionally, whilst a negative relationship between implied volatility and returns is well documented in the papers written by Bekiros et al (2017), Thakolsri, Sethapramote and Jiranyakul (2016) and Dennis et al (2006) we however discovered a positive relationship.

We found a positive relationship between implied volatility and returns and between the change in implied volatility and returns since their slope coefficients were positive in all regressions. This is more in line with Amman, Verhofen and Süss (2008) as well as Rubbaniy et al (2014). Although these papers discovered a significant positive relationship, we discovered a positive but an insignificant relationship. The main finding of this thesis regarding the prediction power of implied volatility is therefore that implied volatility, as well as the change in implied volatility, are not significant predictors of future realised returns. Furthermore, the adjusted $R^2$from our regressions are similar to what has been reported in previous studies where a common factor uniting all is a very low adjusted R^2 which entails that implied volatility generally has very low explanatory power of returns. However, one of our findings is that the adjusted $R^2$for univariate regression using implied volatility as the independent variable (Table 4.5) is negative. This insinuates that implied volatility has nearly no explanatory power of returns which deviates from other studies. Whilst other studies have reported very low adjusted $R^2$for implied volatility these nevertheless imply some explanatory power, whether the estimates are significant or not. Our findings imply some explanatory power of the change in implied volatility but nearly none of implied volatility. The multiple regression with implied volatility and changes in implied volatility as independent variables had a lower adjusted $R^2$of 1% (Table 4.7) than the univariate regression with change in implied volatility as a single independent variable with an $R^2$of 0.8% (Table 4.6). This further supports that implied volatility is a poor predictor of realised returns with it deteriorating the regression model as it lowers the $R^2$when used together with change in implied volatility in a multiple regression.

Even though chosen methodology and data collection is based on previous peer-reviewed research and we have avoided using overlapping data, as suggested by Christensen and Prabhala (1998), there are still risks for errors in our study. One factor that could have affected our results is the fact that we did not winsorize our data to make the residuals of our regression normally distributed, which is an OLS-assumption. We decided to rely on the Central Limit Theorem that samples over

30 observations can be assumed to have a normal distribution, as discussed in section 3.4 of the thesis. However, the fact that we decided to not winsorize our data and that regression model I, II, and III have non-normally distributed error terms could have had an effect on our regressions and made our results misleading. Another factor that could have affected our results associated with realised volatility is the evidence of (positive) autocorrelation present in model I. As positive autocorrelation would cause OLS to understate the true variability of the errors it would increase the risk of rejecting a null hypothesis when it is in fact correct (Brooks, 2008). However, due to time constraints we did not correct for this through a so-called Generalized Least Squares (GLS) method and instead chose to address the risk of it affecting the reliability of the inferences made based upon this model. Other factors that furthermore could explain that parts of our results deviate from the common consensus from previous studies are: usage of different underlying option assets, option styles, sampling methods and sizes, time periods and different empirical methodologies among others.

6. Conclusion

The aim of this study is to examine our two main research questions: 1) If and how well does implied volatility derived from OMXS30 options predict realised index volatility? 2) If and how well does implied volatility derived from OMXS30 options predict realised index returns? These findings are put in relation to how historical volatility can predict realised volatility and how changes in implied volatility can predict returns. First, we will investigate predictability of realised volatility with how implied volatility and historical volatility perform respectively in separate univariate regressions and later together as independent variables in a multivariate regression. Secondly, an predictability investigation will be made for realised returns on how implied volatility and change in implied volatility perform respectively in separate univariate regressions and later together in a multivariate regression. These investigations of predictability of realised volatility and realised returns are made for the period May 2012 to February 2020.

From our study we can conclude that implied volatility is a statistically significant estimate of realised volatility. However, historical volatility is a statistically significant estimate as well that predicts realised volatility with higher accuracy than implied volatility. When implied volatility is used in a multiple regression, implied volatility is no longer a statistically significant estimator while historical volatility is. The multiple regression showed an adjusted R^2 of 12.8% which is higher than the adjusted R^2 of both univariate regressions. Hence, a multiple regression with both implied volatility and historical volatility explains future realised volatility better than if the variables are used in separate univariate regressions. Secondly, our study can conclude that the prediction power of implied volatility and change in implied volatility are both weak and therefore not significant predictors of returns. Implied volatility has a slightly negative prediction power of realised returns with an adjusted R^2 of -0.5% and change in implied volatility has a slight positive prediction power of realised returns with an adjusted R^2 of 0.8%. The prediction power of a multiple regression with implied volatility and change in implied volatility had an adjusted R^2 of 0.1%. Hence, a univariate regression with change in implied volatility can explain future realised returns better than implied volatility and a multiple regression with both as independent variables.

The findings from this thesis are contrary to part of the papers that were studied in the literature survey section of the thesis. Particularly, our finding concerning implied volatility as a predictor of future realised volatility deviates from the results of the majority of previous studies. Our study questions the assumption that implied volatility outperforms historical volatility in predicting future realised volatility. Our deviating results could be a result of our study adopting a different time period and a different market, the Swedish OMXS30 index, compared to previous studies that mainly are concerned with the American S&P 100 index. S&P 100 options are more actively traded than OMXS30 index options and therefore S&P 100 option implied volatility contains more information. This could be a factor that explains our contrary results to previous studies. Our results concerning that implied volatility and change in implied volatility are poor predictors of realised returns are in agreement with most previous studies. Additionally, our study concluded that change in implied volatility has a slightly higher prediction power over implied volatility. Our knowledge contribution to the field of study is that implied volatility is not superior to historical volatility in predicting realised volatility for all markets and time periods and confirming that implied volatility and change in implied volatility are poor predictors of return.

The results of our study could be validated and more reliable from testing of several sub-periods and markets. This could reveal if our conclusions are consistent over smaller sub-periods and other markets, such as the S&P 100 index that most previous studies are based on. A suggestion for further research is to extend this study for a longer time period, adopt different sub-sample periods and compare results for different markets such as different indexes or stocks. Additionally, our study did not cover the concept of the implied volatility spread, which is the difference of the call and put option implied volatility for a certain option. There have been recent studies, such as Han, Bing and Li, Gang (2017), that have discovered that the implied volatility spread can act as a predictor for future returns. Hence, studying this implied volatility spread could be of interest in further studies.

7. Limitations of Research

This thesis investigates the abilities of implied volatility, derived from OMXS30 options, to predict future realised volatility and returns on the OMXS30 index. All of the data used in this study was collected from Thomson Reuters Datastream. The implied volatility data for OMXS30 was available from 10th of May 2012, and we were therefore limited by a specific starting date with our study. Hence, we could not use implied volatility data for a time period earlier than 2012 if we were to use Datastream data. An alternative to using Datastream data is to calculate implied volatilities on our own using the Black-Scholes-Merton model. However, this alternative was not attempted since it would be a too time-consuming process for this thesis. Hence, our study was limited to a specific starting date.

For our method we chose to use the whole period from 10th of May 2012 to 9th of February 2020 for our regressions. In hindsight, more information could perhaps have been harvested from dividing our sample into several sub-samples representing different years. However, this method would be more time consuming and the total time period studied would therefore potentially need to be shortened. A final limitation of our study is that it only attends to data for one specific stock index, OMXS30, and one market, Sweden. Our study has the potential to yield more information and contribute more to the field of study if our results were compared to implied volatility data for other asset types, such as different stocks, or an index from another country, such as the American S&P 100 index. Studying the American S&P 100 index would be an especially interesting addition to this study since it could be compared more fairly to the majority of previous studies, which are based on S&P 100 index options. The decision to not add another asset type or market in our study was due to time constraints and word count limitations.

8. References

8.1 Articles

Amman, M., Verhofen, M. & Süss, S. (2008). Do implied volatility predict stock returns?. *Journal of Asset Management*. 10(4): 222-234.

Bae, J., Kim, C. & Nelson, C.R. (2006). Why are stock returns and volatility negatively correlated?. *Journal of Empirical finance*. 14: 41-58.

Bekiros, S., Jlassi, M., Naoui, K. & Uddin, G.S. (2017). The asymmetric relationship between returns and implied volatility: Evidence from global stock markets. *Journal of Financial Stability*. 30: 156-174.

Bing, H., Gang, L. (2017). Aggregate Implied Volatility Spread and Stock Market Returns. *Rotman School of Management.* Working Paper No. 3047528; Asian Finance Association (AsianFA) 2018 Conference.

Canina, L., Figlewski, S. (1993). The Informational Content of Implied Volatility. *Review of Financial Studies*. 6(3): 659-81.

Christensen, B., Hansen, C. (2002). New evidence on the implied-realized volatility relation. *The European Journal of Finance*. 8(2): 187-205.

Christensen, B., Prabhala, N. (1998). The relation between implied and realized volatility. *Journal of Financial Economics*. 50: 125-150.

Day, T., Lewis, C. (1992). Stock Market Volatility and the Information Content of Stock Index Options. *Journal of Econometrics*. 52(1-2): 267-287.

Dennis, P., Mayhew, S & Stivers, C. (2006). Stock Returns, Implied Volatility Innovations, and the Asymmetric Volatility Phenomenon. *Journal of Financial and Quantitative Analysis*. 41(2): 381-406.

Duan, J-P., Wei, J. (2008). Systematic Risk and the Price Structure of Individual Equity Options, *The Review of Financial Studies*, 22 (5): 1981–2006.

Egbers, T., Swinkels, L. (2015). Can implied volatility predict returns on the carry trade?. *Journal of Banking and Finance*. 59: 14-26.

Harrison, M.J. (1975). The Power of the Durbin-Watson and Geary Tests: Comment and Further Evidence. *The Review of Economics and Statistics. 57*(3): 377-379.

Jiang, G.J., Tian, Y.S. (2005). The Model-Free Implied Volatility and Its Information Content. *The Review of Financial Studies*. 18(4): 1305-1342.

Poon, S., & Granger, C. W. J. (2003). Forecasting Volatility in Financial Markets: A Review. *Journal of Economic Literature*. 41(2): 478-539.

Rubbaniy, G., Asmerom, R., Rizvi, S.K.A & Naqvi, Bushra. (2014). Do fear indices help predict stock returns?. *Quantitative Finance*. 14(5): 831-847.

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.* 27(11), 2151–2175.

Thakolsri, S., Sethapramote, Y. & Jiranyakul, K. (2016). Relationship of the Change in Implied Volatility with the Underlying Equity Index Return in Thailand. *Economic Research Guardian*. 6(2): 74-86.

Yin, Y., Moffatt, P. G. (2019). Correcting the Bias in the Practitioner Black-Scholes Method. *Journal of Risk and Financial Management*. 12(157): 1-12

Zhang, D., Nault, B. R. (2019). Measuring Risk from IT Initiatives Using Implied Volatility. *Journal of the Association for Information Systems*. 20(11): 1566-1588.

8.2 Literature

Brooks, C. (2008). *Introductory Econometrics for Finance*. 2nd edn. New York: Cambridge University Press.

Bryman, A., Bell, E. (2011). *Business Research Methods*. 3rd edn. New York: Oxford University Press.

Davidson, A., Levin, A. (2014). *Mortgage Valuation Models : Embedded Options, Risk, and Uncertainty*. New York: Oxford University Press.

Gunn, M. (2009). T*rading Regime Analysis: The Probability of Volatility Murray Gunn*. Chichester: John Wiley & Sons Ltd.

Hull, J.C. (2017). *Fundamentals of Futures and Options Markets*. 8th edn. Harlow: Pearson Education Limited.

Nations, S. (2012). *Options Math for Traders: How to Pick the Best Option Strategies for Your Market Outlook.* New York: John Wiley & Sons Inc.

Stowell, D. (2012). *Investment Banks, Hedge Funds, and Private Equity*. 2nd edn. Oxford: Academic Press.

8.3 Electronic Resources

Nasdaq Nordic. (n.d.). *Index Weekly Options - the benefits of leverage without overpaying for time value*. Available at: https://www.nasdaq.com/solutions/index-weekly-options-for-omxs30 (Accessed 9 May 2020)

Nasdaq Nordic. (2020). *OMX Stockholm 30 Index (OMXS30)*. Available at: https://indexes.nasdaqomx.com/Index/Overview/OMXS30 (Accessed 9 May 2020)

9. Appendix

Exhibit 1: Regression Results for Model I

Model I: $lnRV_t = \alpha_0 + \beta_1 lnIV_{t-1} + \varepsilon_t$

Dependent Variable: LN_REALISED_VOLATILITY
Method: Least Squares
Date: 05/19/20 Time: 11:16
Sample: 1 92
Included observations: 92

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-1.082909	0.309043	-3.504066	0.0007
LN_IMPLIED_VOLATILITY_LAGED	0.506835	0.171000	2.963958	0.0039

R-squared	0.088931	Mean dependent var	-1.991887
Adjusted R-squared	0.078808	S.D. dependent var	0.381435
S.E. of regression	0.366096	Akaike info criterion	0.849659
Sum squared resid	12.06239	Schwarz criterion	0.904481
Log likelihood	-37.08432	Hannan-Quinn criter.	0.871785
F-statistic	8.785046	Durbin-Watson stat	1.645858
Prob(F-statistic)	0.003886		

Exhibit 1: Regression output for regression model I with realised volatility as the dependent variable and implied volatility as the independent variable.

Exhibit 1.1: White's Test for Heteroskedasticity of Model I

Heteroskedasticity Test: White
Null hypothesis: Homoskedasticity

F-statistic	0.267351	Prob. F(1,90)	0.6064
Obs*R-squared	0.272483	Prob. Chi-Square(1)	0.6017
Scaled explained SS	0.575491	Prob. Chi-Square(1)	0.4481

Test Equation:
Dependent Variable: RESID^2
Method: Least Squares
Date: 05/25/20 Time: 12:19
Sample: 1 92
Included observations: 92

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.194743	0.126430	1.540318	0.1270
LN_IMPLIED_VOLATILITY_LAGED^2	-0.019481	0.037677	-0.517060	0.6064

R-squared	0.002962	Mean dependent var	0.131113
Adjusted R-squared	-0.008116	S.D. dependent var	0.276967
S.E. of regression	0.278089	Akaike info criterion	0.299749
Sum squared resid	6.960019	Schwarz criterion	0.354571
Log likelihood	-11.78845	Hannan-Quinn criter.	0.321875
F-statistic	0.267351	Durbin-Watson stat	1.875514
Prob(F-statistic)	0.606383		

Exhibit 1.1: White's heteroskedasticity test for regression model I.

Exhibit 1.2: Jarque-Bera Normality Test of Regression Residuals in Model I

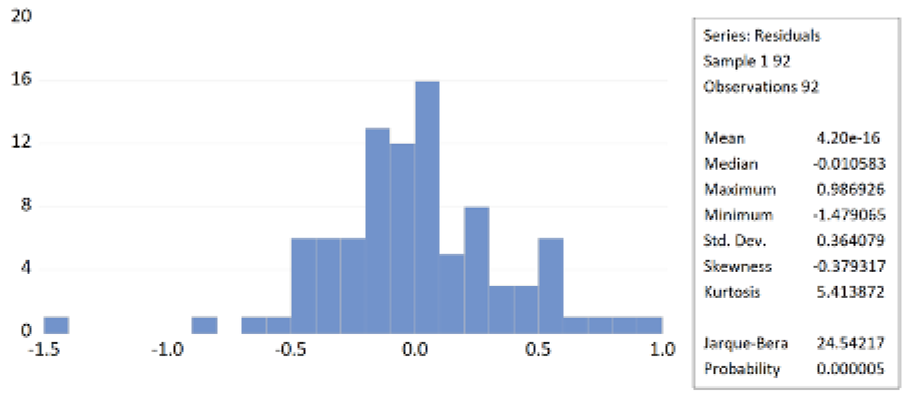

Exhibit 1.2: Jarque-Bera Normality Test for regression model I.

Exhibit 2: Regression Results for Model II

Model II: $lnRV_t = \alpha_0 + \beta_1 lnRV_{t-1} + \varepsilon_t$

Dependent Variable: LN_REALISED_VOLATILITY
Method: Least Squares
Date: 05/19/20 Time: 11:43
Sample: 1 92
Included observations: 92

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-1.362304	0.197704	-6.890607	0.0000
LN_HISTORICAL_VOLATILITY_LAGED	0.317181	0.097761	3.244437	0.0017

R-squared	0.104713	Mean dependent var	-1.991887
Adjusted R-squared	0.094765	S.D. dependent var	0.381435
S.E. of regression	0.362912	Akaike info criterion	0.832185
Sum squared resid	11.85344	Schwarz criterion	0.887007
Log likelihood	-36.28052	Hannan-Quinn criter.	0.854312
F-statistic	10.52637	Durbin-Watson stat	2.214661
Prob(F-statistic)	0.001653		

Exhibit 2: Regression output for regression model II with realised volatility as the dependent variable and historical volatility as the independent variable.

Exhibit 2.1 White's Test for Heteroskedasticity of Model II

Heteroskedasticity Test: White
Null hypothesis: Homoskedasticity

F-statistic	0.325498	Prob. F(1,90)	0.5697
Obs*R-squared	0.331532	Prob. Chi-Square(1)	0.5648
Scaled explained SS	0.980413	Prob. Chi-Square(1)	0.3221

Test Equation:
Dependent Variable: RESID^2
Method: Least Squares
Date: 05/25/20 Time: 12:20
Sample: 1 92
Included observations: 92

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.079597	0.092662	0.859001	0.3926
LN_HISTORICAL_VOLATILITY_LAGE...	0.012041	0.021105	0.570524	0.5697

R-squared	0.003604	Mean dependent var	0.128842
Adjusted R-squared	-0.007467	S.D. dependent var	0.322056
S.E. of regression	0.323257	Akaike info criterion	0.600758
Sum squared resid	9.404532	Schwarz criterion	0.655580
Log likelihood	-25.63489	Hannan-Quinn criter.	0.622885
F-statistic	0.325498	Durbin-Watson stat	1.594673
Prob(F-statistic)	0.569744		

Exhibit 2.1: White's heteroskedasticity test for regression model II.

Exhibit 2.2: Jarque-Bera Normality Test of Regression Residuals in Model II

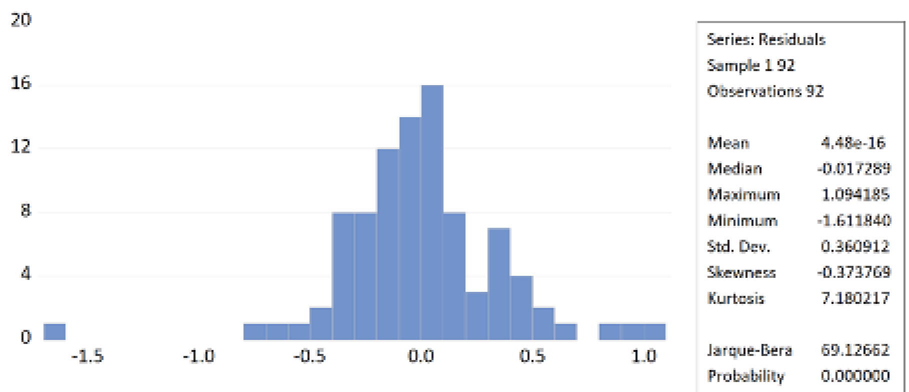

Exhibit 2.2: Jarque-Bera Normality Test for regression model II.

Exhibit 3: Regression Results for Model III

Model III: $lnRV_t = \alpha_0 + \beta_1 lnIV_{t-1} + \beta_2 lnRV_{t-1} + \varepsilon_t$

Dependent Variable: LN_REALISED_VOLATILITY
Method: Least Squares
Date: 05/19/20 Time: 13:04
Sample: 1 92
Included observations: 92

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-1.002260	0.306817	-3.266640	0.0015
LN_IMPLIED_VOLATILITY_LAGED	0.301536	0.197513	1.526665	0.1304
LN_HISTORICAL_VOLATILITY_LAGED	0.226123	0.113910	1.985102	0.0502

R-squared	0.127560	Mean dependent var	-1.991887
Adjusted R-squared	0.107954	S.D. dependent var	0.381435
S.E. of regression	0.360258	Akaike info criterion	0.828074
Sum squared resid	11.55095	Schwarz criterion	0.910306
Log likelihood	-35.09139	Hannan-Quinn criter.	0.861263
F-statistic	6.506357	Durbin-Watson stat	2.126653
Prob(F-statistic)	0.002305		

Exhibit 3: Regression output for regression model III with realised volatility as the dependent variable and implied volatility and historical volatility as the independent variables.

Exhibit 3.1 White's Test for Heteroskedasticity of Model III

Heteroskedasticity Test: White
Null hypothesis: Homoskedasticity

F-statistic	0.176527	Prob. F(2,89)	0.8385
Obs*R-squared	0.363514	Prob. Chi-Square(2)	0.8338
Scaled explained SS	0.927459	Prob. Chi-Square(2)	0.6289

Test Equation:
Dependent Variable: RESID^2
Method: Least Squares
Date: 05/25/20 Time: 12:22
Sample: 1 92
Included observations: 92

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.116480	0.136643	0.852435	0.3963
LN_IMPLIED_VOLATILITY_LAGED^2	-0.013616	0.045749	-0.297623	0.7667
LN_HISTORICAL_VOLATILITY_LAGE...	0.013093	0.022047	0.593884	0.5541

R-squared	0.003951	Mean dependent var	0.125554
Adjusted R-squared	-0.018432	S.D. dependent var	0.294783
S.E. of regression	0.297488	Akaike info criterion	0.445177
Sum squared resid	7.876401	Schwarz criterion	0.527409
Log likelihood	-17.47814	Hannan-Quinn criter.	0.478367
F-statistic	0.176527	Durbin-Watson stat	1.649991
Prob(F-statistic)	0.838468		

Exhibit 3.1: White's heteroskedasticity test for regression model III.

Exhibit 3.2: Jarque-Bera Normality Test of Regression Residuals in Model III

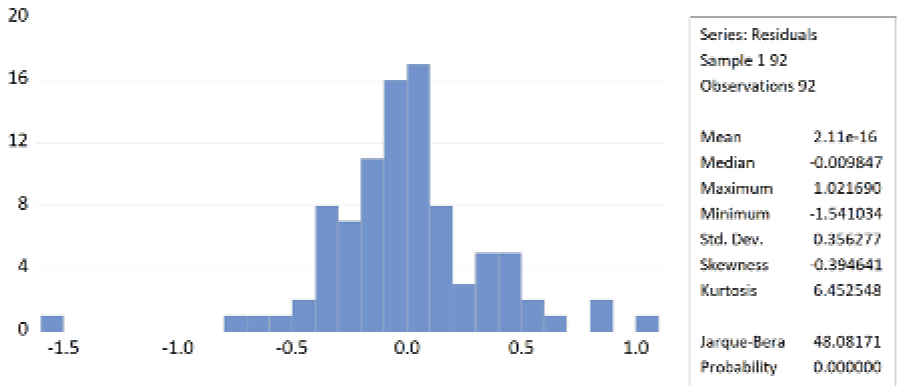

Exhibit 3.2: Jarque-Bera Normality test for regression model III.

Exhibit 3.3: Variance Inflation Factors Test for Multicollinearity in Model III

Variance Inflation Factors
Date: 05/19/20 Time: 13:07
Sample: 1 92
Included observations: 92

Variable	Coefficient Variance	Uncentered VIF	Centered VIF
C	0.094137	66.72955	NA
LN_IMPLIED_VOLATI...	0.039011	90.32341	1.377731
LN_HISTORICAL_VO...	0.012976	37.61691	1.377731

Exhibit 3.3: Variance Inflation Factors test for regression model III.

Exhibit 4: Regression Results for Model IV

Model IV: $r_t = \alpha_0 + \beta_1 lnIV_{t-1} + \varepsilon_t$

Dependent Variable: RETURNS
Method: Least Squares
Date: 05/25/20 Time: 12:24
Sample: 1 92
Included observations: 92

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.001405	0.001524	0.921822	0.3591
LN_IMPLIED_VOLATILITY_LAGED	0.000611	0.000843	0.725291	0.4702

R-squared	0.005811	Mean dependent var	0.000308
Adjusted R-squared	-0.005236	S.D. dependent var	0.001800
S.E. of regression	0.001805	Akaike info criterion	-9.775050
Sum squared resid	0.000293	Schwarz criterion	-9.720229
Log likelihood	451.6523	Hannan-Quinn criter.	-9.752924
F-statistic	0.526047	Durbin-Watson stat	2.283891
Prob(F-statistic)	0.470155		

Exhibit 4: Regression output for regression model IV with realised returns as the dependent variable and implied volatility as the independent variable.

Exhibit 4.1 White's Test for Heteroskedasticity for Model IV

Heteroskedasticity Test: White
Null hypothesis: Homoskedasticity

F-statistic	0.003288	Prob. F(1,90)	0.9544
Obs*R-squared	0.003361	Prob. Chi-Square(1)	0.9538
Scaled explained SS	0.003924	Prob. Chi-Square(1)	0.9501

Test Equation:
Dependent Variable: RESID^2
Method: Least Squares
Date: 05/25/20 Time: 12:24
Sample: 1 92
Included observations: 92

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	3.31E-06	2.29E-06	1.448532	0.1509
LN_IMPLIED_VOLATILITY_LAGED^2	-3.91E-08	6.82E-07	-0.057342	0.9544

R-squared	0.000037	Mean dependent var	3.19E-06
Adjusted R-squared	-0.011074	S.D. dependent var	5.01E-06
S.E. of regression	5.03E-06	Akaike info criterion	-21.53945
Sum squared resid	2.28E-09	Schwarz criterion	-21.48463
Log likelihood	992.8147	Hannan-Quinn criter.	-21.51732
F-statistic	0.003288	Durbin-Watson stat	1.315904
Prob(F-statistic)	0.954400		

Exhibit 4.1: White's heteroskedasticity test for regression model IV.

Exhibit 4.2: Jarque-Bera Normality Test of Regression Residuals in Model IV

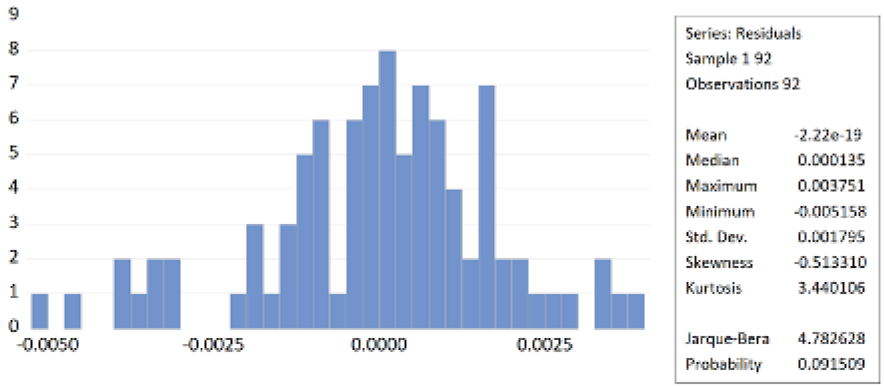

Exhibit 4.2: Jarque-Bera Normality Test for regression model IV.

<u>Exhibit 5: Regression Results for Model V</u>

Model V: $\boldsymbol{r_t = \alpha_0 + \beta_1 \Delta IV_{t-1} + \varepsilon_t}$

Dependent Variable: RETURNS
Method: Least Squares
Date: 05/25/20 Time: 12:28
Sample: 1 92
Included observations: 92

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.000314	0.000187	1.678395	0.0967
DELTA_LN_IMPLIED_VOLATILITY_LAGED	0.016957	0.012778	1.327045	0.1879

R-squared	0.019192	Mean dependent var	0.000308
Adjusted R-squared	0.008294	S.D. dependent var	0.001800
S.E. of regression	0.001793	Akaike info criterion	-9.788600
Sum squared resid	0.000289	Schwarz criterion	-9.733779
Log likelihood	452.2756	Hannan-Quinn criter.	-9.766474
F-statistic	1.761048	Durbin-Watson stat	2.199043
Prob(F-statistic)	0.187851		

Exhibit 5: Regression output for regression model V with realised returns as the dependent variable and the change in implied volatility as the independent variable.

Exhibit 5.1 White's Test for Heteroskedasticity in Model V

Heteroskedasticity Test: White
Null hypothesis: Homoskedasticity

F-statistic	0.066878	Prob. F(1,90)	0.7965
Obs*R-squared	0.068313	Prob. Chi-Square(1)	0.7938
Scaled explained SS	0.070581	Prob. Chi-Square(1)	0.7905

Test Equation:
Dependent Variable: RESID^2
Method: Least Squares
Date: 05/25/20 Time: 12:29
Sample: 1 92
Included observations: 92

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	3.19E-06	5.18E-07	6.155516	0.0000
DELTA_LN_IMPLIED_VOLATILITY_LAGE...	-0.000215	0.000830	-0.258607	0.7965

R-squared	0.000743	Mean dependent var	3.14E-06
Adjusted R-squared	-0.010360	S.D. dependent var	4.65E-06
S.E. of regression	4.67E-06	Akaike info criterion	-21.68953
Sum squared resid	1.96E-09	Schwarz criterion	-21.63471
Log likelihood	999.7185	Hannan-Quinn criter.	-21.66741
F-statistic	0.066878	Durbin-Watson stat	1.306503
Prob(F-statistic)	0.796529		

Exhibit 5.1: White's heteroskedasticity test for regression model V.

Exhibit 5.2: Jarque-Bera Normality Test of Regression Residuals in Model V

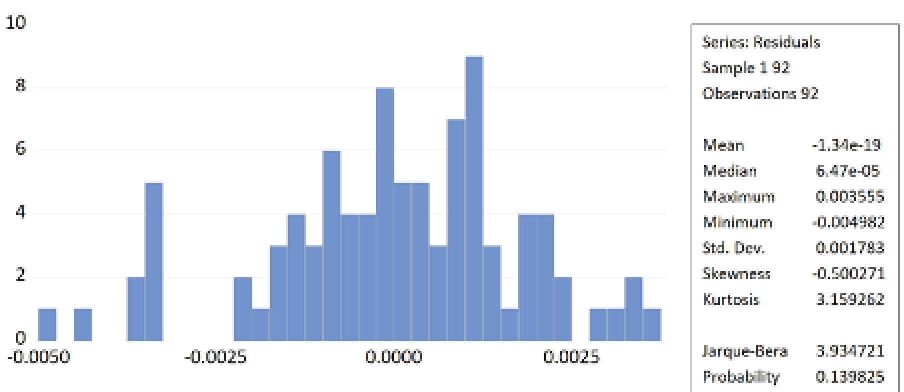

Exhibit 5.2: Jarque-Bera Normality Test for regression model V.

Exhibit 6: Regression Results for Model VI

Model VI: $r_t = \alpha_0 + \beta_1 lnIV_{t-1} + \Delta IV_{t-1} + \varepsilon_t$

Dependent Variable: RETURNS
Method: Least Squares
Date: 05/25/20 Time: 12:31
Sample: 1 92
Included observations: 92

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.001153	0.001533	0.752110	0.4540
LN_IMPLIED_VOLATILITY_LAGED	0.000468	0.000849	0.551552	0.5826
DELTA_LN_IMPLIED_VOLATILITY_LAGED	0.015979	0.012950	1.233913	0.2205

R-squared	0.022533	Mean dependent var	0.000308
Adjusted R-squared	0.000567	S.D. dependent var	0.001800
S.E. of regression	0.001800	Akaike info criterion	-9.770274
Sum squared resid	0.000288	Schwarz criterion	-9.688041
Log likelihood	452.4326	Hannan-Quinn criter.	-9.737084
F-statistic	1.025821	Durbin-Watson stat	2.186332
Prob(F-statistic)	0.362701		

Exhibit 6: Regression output for regression model VI with realised returns as the dependent variable and implied volatility and the change in implied volatility as the independent variables.

Exhibit 6.1 White's Test for Heteroskedasticity in Model VI

Heteroskedasticity Test: White
Null hypothesis: Homoskedasticity

F-statistic	0.037934	Prob. F(2,89)	0.9628
Obs*R-squared	0.078358	Prob. Chi-Square(2)	0.9616
Scaled explained SS	0.080402	Prob. Chi-Square(2)	0.9606

Test Equation:
Dependent Variable: RESID^2
Method: Least Squares
Date: 05/25/20 Time: 12:32
Sample: 1 92
Included observations: 92

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	3.50E-06	2.15E-06	1.628953	0.1069
LN_IMPLIED_VOLATILITY_LAGED^2	-9.98E-08	6.39E-07	-0.156223	0.8762
DELTA_LN_IMPLIED_VOLATILITY_LAGE...	-0.000188	0.000839	-0.223777	0.8234

R-squared	0.000852	Mean dependent var	3.13E-06
Adjusted R-squared	-0.021601	S.D. dependent var	4.67E-06
S.E. of regression	4.72E-06	Akaike info criterion	-21.65929
Sum squared resid	1.98E-09	Schwarz criterion	-21.57705
Log likelihood	999.3272	Hannan-Quinn criter.	-21.62610
F-statistic	0.037934	Durbin-Watson stat	1.310223
Prob(F-statistic)	0.962792		

Exhibit 6.1: White's heteroskedasticity test for regression model VI.

Exhibit 6.2: Jarque-Bera Normality Test of Regression Residuals in Model VI

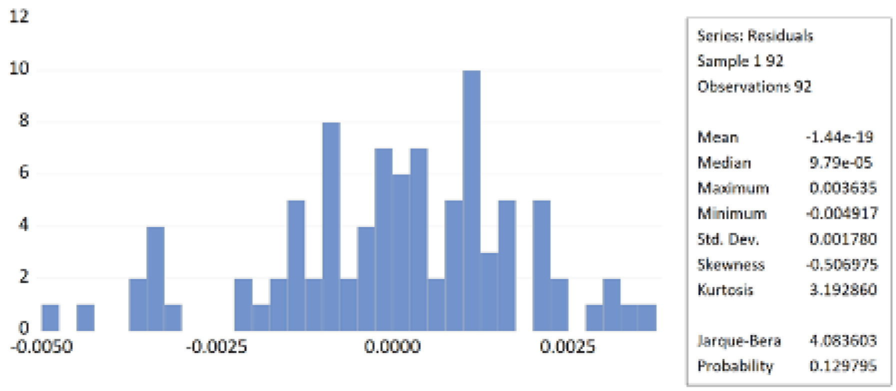

Exhibit 6.2: Jarque-Bera Normality test for regression model VI.

Exhibit 6.3: Variance Inflation Factors Test for Multicollinearity of Model VI

Variance Inflation Factors
Date: 05/19/20 Time: 11:36
Sample: 1 92
Included observations: 92

Variable	Coefficient Variance	Uncentered VIF	Centered VIF
C	2.35E-06	66.74091	NA
LN_IMPLIED_VOLATI...	7.20E-07	66.81269	1.019115
DELTA_IMPLIED_VO...	0.000168	1.019689	1.019115

Exhibit 6.3: Variance Inflation Factors test for regression model VI.

Stockholm Business School
Stockholm University
SE-106 91 Stockholm
Tel: 08 – 16 20 00
www.sbs.su.se

Stockholm University