Zeppelin Universität

Chair of Empirical Finance and Econometrics

The Information Content of VIX Volatility

Humboldt Research Project Project

in

Modul 11251 Forschungsprojekt

Name:	Sophia Charlotte Gläser
Student number:	15202284
Programme:	Corporate Management and Economics
Term:	Fall Semester 2018
Examinor:	Prof. Dr. Franziska Peter
Due date:	31.01.2019

Contents

List of Tables

List of Figures

Acronyms

BS Black-and-Scholes-Merton Model

CBOE Chicago Board of Options Exchange

SPX S&P 500 index

VIX Volatility index

Abstract

This paper investigates the information content of model-free implied volatility for daily realized volatility of the S&P 500, using the Volatility index (VIX) from the Chicago Board of Options Exchange and daily volatilities from 2000 to 2017. In contrast to earlier implied volatilities, the VIX is not based on any specific option pricing model. Therefore, it provides a direct test of market efficiency and does not suffer from the joint hypothesis problem. For the statistical analysis the approach from an HAR-RV model as described by Corsi (2009) is used, thus including not only daily, but also weekly and monthly historic volatility in the encompassing regression analysis. The results show, that the VIX provides additional information compared to historic volatility, but is not able to subsume all the information contained in the historic volatilities, which partly contradicts previous research. This results are robust to serial correlation and alternative sampling methods.

1 Introduction

Financial market volatility is of high interest for the financial sector. It is for example a key input to the pricing of financial instruments like derivatives, or to risk measures such as the Value at Risk. Moreover, it gives information on the risk-return trade-off, which is a central question in portfolio allocation and managerial decision making.

Seeing its importance, it is not astonishing that considerable research has been devoted to the question how volatility, which is not directly observable, can be estimated and predicted. Whereas earlier approaches were mainly reliant on ARCH or stochastic volatility models using historic volatility, there has recently been a growing interest in volatility implied from option price data (Bakanova, 2010).

Since options are contracts giving the holder the right to buy or sell an underlying asset at a specified date in the future, they are said to have a "forward-looking nature", meaning that they are supposed to be highly related to the market's expectation about the future volatility of the underlying asset over the remaining life of the option. Therefore, if markets are efficient and the model pricing the option is specified correctly, the volatility implied from option prices should be an unbiased and efficient estimator of future realized volatility (Bakanova, 2010).

One popular approach for estimating implied volatility is the Black-and-Scholes-Merton Model (BS) implied volatility. The BS model is an option pricing model with volatility as an input factor. Using observed option prices as the input and solving for volatility, it is possible to obtain a volatility measure that is widely believed to be "informationally superior to the historic volatility of the underlying asset" (Jiang & Yisong, 2005, p. 1305). Whereas early studies found this to be a biased forecast, the correction of several methodological problems provided evidence that BS implied volatility contains statistically significant information and is a more efficient forecast of realized volatility than historic volatility (Jiang & Yisong, 2005).

Notwithstanding this results, the BS implied volatility has some specification problems. Firstly, it focuses on at-the-money options. At-the-money options are usually the once most actively traded and thus the most liquid ones, but this fails to include information contained in other options. Moreover, tests based on the BS equation are joined tests of market efficiency and the option pricing model, thus they suffer from model misspecification errors (Jiang & Yisong, 2005).

That is why during the last years, implied volatility indices which are not based on a pricing assumption have gained popularity, starting with the paper by Britten-Jones and Neuberger (2000). Contrary to earlier implied volatility, their approach is derived directly from option prices and the no-arbitrage condition. Showing that the risk-neutral return variance can be derived entirely from option price data, they provide a volatility measure which does not suffer from the joined hypothesis problem and moreover allows to incorporate not only at-the-money options. Several papers, such as Bakanova (2010), Taylor, Yadav and Zhang (2010) or Jiang and Yisong (2005), used the results from Britten-Jones and Neuberger (2000), extended their approach and tested the informational efficiency of the model-free implied volatility. Jiang and Yisong (2005) for example, find that the model-free implied variances over a 30-day horizon subsumes all information contained in both BS and historic variance, using OLS regression. Curiously, conducting the same regression with volatility data, their model-free implied volatility was not significant when historic volatility was included.

In 2003, the Chicago Board of Options Exchange (CBOE) reacted to this overall confirming academic findings and changed the calculation of its Volatility index (VIX) from the BS to the model-free implied volatility (Fleming, Ostdiek & Whaley, 1995). This paper directly uses the VIX from CBOE and tests the data for a 17-year time period, including the financial crisis around 2008, using both univariate and encompassing regression analysis. Moreover, the historic volatility information included in the model is extended. Previous research mainly used only the one-day-lagged realized volatility, whereas this paper uses the approach from the HAR-RV model from Corsi (2009) and includes equally weekly and monthly historic volatility averages in the regression.

The paper is structured as follows. Section 2 introduces the necessary basics about volatility, describing the concept used for realized volatility and modelling it. This is important for the analysis, because it justifies the approach used and allows to draw conclusions from the empirical findings. Section 3 presents the data set and methodology used to investigate the information content of VIX volatility, and is followed by the empirical findings in section 4. Section 5 summarizes and concludes.

2 Selected Estimation and Modelling Procedures for Volatility

Volatility can not be directly observed and thus has to be estimated. This section briefly presents a concept for estimating realized volatility and an approach for modelling it which is able to capture very well the stylized facts observed with financial data.

2.1 Estimating and Modeling Volatility using Historic Volatility

2.1.1 Estimating Realized Volatility

By definition "volatility seeks to capture the strength of the (unexpected) return variation over a given period of time" (Andersen, Bollerslev & Diebold, 2002, p.7).

With volatility being a latent variable, there are multiple concepts for volatility. According to Andersen et al. (2002) they can be grouped in (i) the *notional volatility* corresponding to the ex-post sample-path return variability over a fixed time interval, (ii) the ex-ante *expected volatility* over a fixed time interval or the (iii) the *instantaneous volatility* corresponding to the strength of the volatility process at a point in time. For this paper the ex-post sample-path return variability is of importance.

Volatility measures usually represent the average volatility over a discrete time period, as a continuous record of price data is not available, and even for very liquid markets, price data are distorted by microstructure effects. It can however be shown, that under some assumptions the sum of squared high-frequency returns is a consistent estimator for the return variance. This results mainly build on the work of Andersen et al. (2002) and will be briefly introduced here. As terminology is not consistent in previous research, in this paper this sum of squared high-frequency returns is the return variance, and return volatility its square root, the standard deviation.

To begin with, a standard continuous-time no-arbitrage setting should be invoked. As return volatility aims to capture the strength of the unexpected return variation, the component of a price change as opposed to an expected price movement needs to be defined. Andersen et al. (2002) show that under certain assumptions the instantaneous return process can be decomposed into an expected return component and a martingale innovation (in the discrete time setting this decomposition is more complex, for this paper it shall only be relevant that the martingale part is still the dominant contribution to the return variation over short intervals). The cumulative sample path variability of this martingale component can be represented by a quadratic variation process, and they define the ex-post measured *notional variance*[1] as the increment to this quadratic variation. The authors show that this notional variance can be consistently estimated, using high-frequency returns or a large sample of returns, with the *realized variance*, defined as:

$$v^2(t,h;n) = \sum_{i=1}^{n} r(t-h+(i/n)\cdot h, h/n)^2 \tag{1}$$

over any fixed $[t-h,t], 0<h$ time interval (their definition 6 and proposition 5). This realized variance is simply the second sample moment of the return process, scaled by the number of observations n.

So in summary, the increment to the quadratic return variation, which is the cumulative past return variance, can be consistently and well approximated through the accumulation of squared high-frequency returns. Taking the square root of this realized variance, the *realized volatility* is obtained.

[1]Andersen et al. (2002) term this *notional volatility*, but in alignment with the above definition, it is referred to as notional variance in this paper.

2.1.2 Modelling Volatility - the HAR-RV Model

Having introduced the concept of measuring realized volatility, there are multiple approaches that try to model the volatility process[2]. Although volatility is not directly observable, returns and their volatility have some characteristics that are commonly examined and can be used to build volatility models.

These stylized facts include high excess kurtosis for daily return series and clustering of return variability, meaning that periods of high volatility are followed by high, and periods of low volatility are followed by low volatility (Tsay, 2005). Moreover, the autocorrelations of the square and absolute returns show a very strong persistence over long time periods. If return distributions with different time horizons are observed, they show tail-crossover, i.e. the shape depends on the time scale and convergence to normal distribution is slow. Financial data also show evidence of scaling and multiscaling (Corsi, 2009).

There are various modelling approaches for volatility. This paper will refer to the HAR-RV model, based on Corsi (2009), as it is able to reproduce the stylized facts described above very well. The model presents a volatility cascade from low to high frequencies, in which not only daily lagged realized volatility, but also weekly and monthly volatility influence future realized volatility.

The model shall be briefly introduced. It assumes that prices follow the standard continuous-time process, represented by the stochastic differential equation

$$dp(t) = \mu(t)dt + \sigma(t)dW(t), \ 0 \leq t \leq T \tag{2}$$

where $p(t)$ is the logarithm of the instantaneous price, μ is a finite variation stochastic process, $W(t)$ standard Brownian motion and σ a stochastic process independent of $W(t)$. For this process, the variance is the *integrated variance*,

[2]An extensive overview of the literature on volatility modelling and forecasting can be found in Poon and Granger (2003)

which is integral of the instantaneous variance over the one-day interval

$$IV_t^{(d)} = \int_{t-1d}^{t} \sigma^2(w)dw, \tag{3}$$

and the corresponding volatility is its square root: $\sigma = \sqrt{IV_t^{(d)}}$. Andersen et al. (2002) show that instead of the abstract martingale representation of the return decomposition that was described in section 2.1.1, the continuous sample paths can also be represented using stochastic differential equations, as it is done here. Then the integrated variance equals the notional volatility, and can equally be estimated using the sum of squared returns. Thus, the HAR-RV model uses the following approximation for realized volatility over the one-day interval, which will be adopted in this paper:

$$RV_t^{(d)} = \sqrt{\sum_{j=0}^{M-1} r_{t-j\times\Delta}^2} \tag{4}$$

with $\Delta = 1d/M$ being the sampling frequency and $r_{t-j\times\Delta}^2$ defined as the continuously compounded Δ-frequency returns (with t the day and j the time within the day).

The idea behind the HAR-RV model is closely connected to the *Heterogeneous Market Hypothesis*, which describes the presence of heterogeneity across market participants and was presented by Müller et al. (1993). This view of financial markets grounds on the fractal model, introduced and applied to financial markets by Mandelbrot (1963). The approach of the fractal model is to analyse (time series) objects on different time scales and compare the obtained results. The argument is that conventional time series analysis, focusing on regularly spaced observations, does not capture the real nature of the raw data, as the usual time choice for recording observations (e.g. a day) is arbitrary.

Using this fractal approach and empirical findings of volatility characteristics[3] led to the heterogeneous market hypothesis. This hypothesis states that the market gives rise to heterogeneous trading behaviours since different market participants or components have different time horizons for their trading goals and for their consideration of past events. The time span has on the one end the high-frequency traders such as market makers, in the middle medium term traders and on the other end the low-frequency traders, such as central banks or commercial organizations. Driven by this components, "the market is heterogeneous with a 'fractal' structure of the participants' time horizon" (Müller et al., 1993, p.12).

Corsi (2009) adds to the observations of Müller et al. (1993) that volatility has an asymmetric behaviour of influence, meaning that volatility over longer time periods has a stronger influence on volatility observed over short periods than conversely. The pattern that emerges is a volatility cascade from low to high frequencies.

To formalize the model the *latent partial volatility* $\tilde{\sigma}_t^{(\cdot)}$ is defined as the volatility generated by a certain market component. To account for short-term, medium-term and long-term traders, the time horizons of one day (d), one week (w) and one month (m) are considered, and denoted by $\tilde{\sigma}_t^{(d)}$, $\tilde{\sigma}_t^{(m)}$ and $\tilde{\sigma}_t^{(w)}$. In the model, each of this volatility components corresponds to a market participant that forms the expectation for the volatility of one period ahead based on both the observation of the current realized volatility, according to the own time frame, and on the expectation of the one horizon longer volatility. For example

[3] Müller et al. (1993) observed that the decline of the return autocorrelation function is not exponential, as suggested for example by lower-order GARCH or ARCH models, but rather hyperbolic. Assuming that each of the distinct components has an exponential decline with different time horizons, in sum comes close to a hyperbolic decline. Moreover, if market participants were homogeneous, volatility should be negatively correlated with market activity, as the price should converge to the "real value". However, they are positively correlated, which might be explained by the fact that actors react and execute in different market situations (Müller et al., 1993).

for the market participant with a daily horizon, the latent volatility would be

$$\tilde{\sigma}^{d}_{t+1d} = c^{(d)} + \Phi^{(d)} RV^{(d)}_{t} + \gamma^{(d)} E_t[\tilde{\sigma}^{(w)}_{t+1w}] + \tilde{w}^{(d)}_{t+1d},$$

where $\tilde{w}^{(d)}_{t+1d}$ is the return innovation. Substitution the latent volatilities of the different horizons and defining the latent volatility as the daily integrated volatility ($\sigma^{(d)}_{t}$), the cascade model can be written as

$$\sigma^{(d)}_{t+1d} = c + \beta^{(d)} RV^{(d)}_{t} + \beta^{(w)} RV^{(w)}_{t} + \beta^{(m)} RV^{(m)}_{t} + \tilde{w}^{(d)}_{t+1d}, \tag{5}$$

Observing the volatility data ex-post, $\sigma^{(d)}_{t+1d}$ can also be written as the realized volatility. Thus, a functional form for time series representation is obtained with

$$RV^{(d)}_{t+1d} = c + \beta^{(d)} RV^{(d)}_{t} + \beta^{(w)} RV^{(w)}_{t} + \beta^{(m)} RV^{(m)}_{t} + w_{t+1d}, \tag{6}$$

where w_{t+1d} includes the innovation component and the measurement and estimation error. The realized volatility for the weekly and monthly aggregated periods are simply the rolling averages over the respective periods.

Using simulated return data, Corsi (2009) shows that the HAR-RV absolute returns and volatility reproduce the stylized facts mentioned above very well. The data has not only the excess of kurtosis but also the tail cross-over and is able to reproduce the long memory of the empirical data. Moreover, using OLS and out-of-sample forecast, Corsi (2009) shows that the past volatility components all have statistically significant information content for one day ahead realized volatility and show good out-of-sample forecasting performance.

2.2 Estimating Volatility Using Option Price Data

2.2.1 The General Idea and Evolution of Implied Volatility

Apart from using historic data for volatility modelling, the forward-looking nature of options data can be used to augment the information set. This approach is termed *implied volatility.* The intuition behind implied volatility is that option prices can be seen as reflecting the market participants' expectations about the future movements of the underlying asset. Assuming that the market is efficient, as described by Malkiel and Fama (1970), and that the asset pricing model is correct, the implied volatility derived from the option prices should not only subsume all information contained in historic volatility but also be a more efficient forecast of future volatility (Jiang & Yisong, 2005).

One example for implied volatility is the *BS implied volatility*, building on the asset pricing model as presented by Black and Scholes (1973), which uses volatility as input for pricing options. Having the option prices available on the market, it is possible to extract a value for the expected volatility by inverting the theoretical asset pricing model.

Whereas previous research yielded ambiguous results concerning the informational efficiency and forecasting ability of BS implied volatility, recent research made several methodological and data corrections and found evidence supporting the hypothesis that implied volatility has predictive power for future volatility. These corrections included using longer time series, high-frequency asset returns or non-overlapping samples (Jiang & Yisong, 2005).

However, there are several disadvantages of implied volatility from an asset pricing model. One drawback is that the BS implied volatility relies mostly on information from at-the-money options, which are generally the most actively traded ones. This fails to incorporate information contained in other options (such as out-of-the money options). Moreover and more importantly, testing for the BS implied volatility always means testing for market efficiency and the BS model with its assumptions (e.g. constant volatility, no transaction costs, no arbitrage, continuous trading, constant risk-free interest rate, divisible securities and no short sell (Poon & Granger, 2003)) jointly. As it is not possible in this setting to test for market efficiency or the pricing assumptions separately, these tests are subject to model misspecification errors (Jiang & Yisong, 2005). Because of this rather critical assumptions and the resulting flaws of implied volatility obtained from the BS model, an implied volatility that does not rely on an asset pricing model was introduced.

2.2.2 Model-Free Implied Volatility - the VIX

The idea of model free implied volatility was presented by Britten-Jones and Neuberger (2000) who show that the risk-neutral realized volatiltiy can be derived from a set of options with matching expirations, thus extending the approach of Derman and Kani (1994), Dupire (1994), Dupire (1997) and Rubinstein (1994) on implied distributions. Contrary to the BS implied volatility, no assumptions are made about the pricing process. Instead, a complete set of option prices is taken as given and as much information as possible is extracted about the underyling price process (Britten-Jones & Neuberger, 2000).

Britten-Jones and Neuberger (2000) show in an approach resembling a binomial tree that the probability of the stock price reaching any particular level and of a price move is determined by the initial set of option prices. Moreover, they show that the risk-neutral expected sum of squared returns between two dates is given from the set of options expiring on these two dates (their proposition 2). This formula does not use an option pricing model and is only based on no-arbitrage conditions, therefore it solves the joint-hypothesis problem and tests directly for market efficiency (Jiang & Yisong, 2005).

Several paper used this result and compared the informational efficiency of model-free implied, BS and historic volatility. One example is Jiang and Yisong (2005). They use the approach of Britten-Jones and Neuberger (2000), extend the formula to asset price processes with jumps and test the informational efficiency of this model-free implied volatility in comparison to both BS implied and historic volatility, using regression analysis. With options data from the S&P 500 index (SPX) and 5-min return data, they examine a 6-year sample period between 1988 and 1994. They find that the model-free implied variance subsumes the information contained in the BS implied and past realized variance. Another example is the paper of Bakanova (2010). With daily data from crude oil futures between November 1986 and December 2006, he evaluates the information content of model-free implied volatility in comparison to historic volatility. Using regression analysis, he comes to the result that implied volatility subsumes the information contained in historical volatility. In contrast to these results is for example Taylor et al. (2010). With individual stock data from 149 U.S. firms between January 1996 and December 1999, they find that for a one-day-ahead estimation, historic volatility outperforms model-free implied volatility. When however extending the prediction horizon or using the most actively traded options, the option implied volatility is more informative for most of the stocks.

One of the first implemented implied volatility indices is the VIX from CBOE. It is computed each trading day on a real-time basis, with data available dating back to January 1986. Its introduction in 1993 had the intention to provide both a benchmark for market volatility in the short term, and a volatility index on which futures and options could be written and traded.

At this time the VIX was based on the BS option pricing model using S&P 100 options, as they where the most actively traded in the U.S. (Fleming et al., 1995). In 2003 the VIX adopted to the changes that took place in the options market. First, the SPX option market superseded the S&P 100 options market as the most actively traded option market in the U.S., thus following 2003 the VIX was based on SPX options. Secondly, option index trading behaviour changed. Whereas in the 1990s both call and put index options where equally important, over the years both out-of-the money and at-the-money puts gained popularity since they were increasingly bought by portfolio insurers. Thus the VIX started to include out-of-the money options in its calculation, capturing the increasing demand in put options and bringing another advantage compared to BS implied volatility (Whaley, 2008). Finally, the VIX calculation changed in 2003, adopting to the model-free implied volatility approach, which was in 2003 widely used by financial theorists, risk managers, and volatility traders alike (Chicago Board of Options Exchange, 2009).

The VIX is constructed in the way that it eliminates mis-specification and "smile" effects thus making it an accurate measurement of implied volatility (Blair, Poon & Taylor, 2001). The formula for its calculation is

$$\sigma^2 = \frac{2}{T}\sum_i \frac{\Delta K_i}{K_i^2} e^{RT} Q(K_i) - \frac{1}{T}(\frac{F}{K_0} - 1)^2 \tag{7}$$

$$VIX = 100 \times \sqrt{\{T_1\sigma_1^2[\frac{N_{T_2} - N_{30}}{N_{t_2} - N_{T_1}}] + T_2\sigma_2^2[\frac{N_{30} - N_{T_1}}{N_{t_2} - N_{T_1}}]\} \times \frac{N_{365}}{N_{30}}} \tag{8}$$

with T the time to expiration, F the forward index level, K_0 the first strike below the forward index level, K_i the strike of the i^{th} out-of-the-money option, ΔK_i the interval between the strike prices, R the risk-free rate, $Q(K_i)$ the midpoint of bid-ask spread for each option with strike K_i and N the number of minutes to expiration (Chicago Board of Options Exchange, 2009).

The VIX always uses near- and next-term put and call options, with more than 23 and less than 37 days to expiration. Each week, the options used for calculation roll over to new maturities, thus long-term options become next-term options, and in the following week fall out of the sample. Hence the VIX will always reflect an interpolation of the volatility between these two option maturities (Poon & Granger, 2003). After the selection of the options for the calculation, the volatility for both the near- and next-term options is calculated, using equation 7. Out of this two volatilities, the 30-day-weighted average is calculated (equation 8) (Chicago Board of Options Exchange, 2009).

3 Methodology and Data

The aim of this paper is to build on the results from previous research, for example Jiang and Yisong (2005), who use their results to justify both the general validity of the model-free implied volatility and the decision of CBOE to modify the calculation of the VIX. This paper will examine directly the information content of the VIX compared to historic volatility, using a more recent and longer time period of 17 years, between January 2000 and December 2017. Moreover, more information regarding the historic volatility is included with the use of the HAR-RV model by Corsi (2009). In this section, first the sample and the model is introduced, then the research questions and testable hypothesis are formalized. Finally the limitations of the methodology are briefly discussed.

3.1 Data and Calculation of Input Factors

The data used in this study is obtained from several sources. The realized variance of the SPX is from the 'Oxford-Man Institute's realized library V0.2' (2009). Daily VIX values are the closing values from CBOE. Since September 22, 2003 CBOE started to use the model-free implied volatility for the VIX, but calculated the prices with the new methodology ex-post, dating back until 1990. As the sample from this study covers both the period before and after the revised methodology, the data was taken both from the ex-post Chicago Board of Options Exchange (2018b) and the daily-updated Chicago Board of Options Exchange (2018c). The SPX closing values, used only for visualisation, are also taken from Chicago Board of Options Exchange (2018a). The sample consists of daily data from January 2000 to December 2017.

The calculation used for the daily realized variance from the 'Oxford-Man Institute's realized library V0.2' (2009) is the sum of squared 5-min high-frequency returns, as explained in section 2.1.1. Their formula is given by

$$\sigma_t = \sum x_t^2 \tag{9}$$

with $x_t = X_t - X_{t-1}$ and X_t the logarithm of the price at time t.

Previous research has used different sampling frequencies for the calculation of realized volatility. Whereas earlier studies used mainly daily returns, more recent studies argue in favour of using intraday returns, as they have certain advantages over daily data. For example Andersen, Bollerslev, Diebold and Labys (2003) point out that high-frequency returns help both for predicting again high-frequency returns, but also that they contain information for longer horizons, such as monthly or quarterly. Moreover Andersen and Bollerslev (1998) show that realized volatility calculated using squared returns produces inaccuracies with daily returns.

The data obtained was in a cleaned format. First, entries outside the timestamp when exchanges are open had been deleted. Secondly, entries with the same time stamp had been replaced with the median bid-ask price. Thirdly, entries with a negative spread (as they violate the no-arbitrage condition) or extremely large spread (50 times the median of the day) had been deleted. Lastly, entries for which the mid-quote deviated largely from the mean had also been removed.

For this paper, to obtain the realized volatility, the square root of the realized variance is taken. For a more intuitively comparison to the VIX, the values are multiplied by 100:

$$RV_t = \sqrt{\sum x_t^2} \times 100. \tag{10}$$

For the historic volatility the lagged realized volatility is used. As the approach from the HAR-RV model presented in 2.1.2 is integrated, additionally to lagged daily realized volatility, the lagged weekly and lagged monthly volatility is needed. They are computed using the rolling average over the respective time periods, thus weekly realized volatility over the last 5 trading days is calculated as

$$RV_t^{(w)} = \frac{1}{5}(RV_t^{(d)} + RV_{t-1d}^{(d)} + ... + RV_{t-4d}^{(d)}). \tag{11}$$

The monthly volatility is calculated alike, using a period of 20 days.
For the VIX the daily closing value is taken, as it contains the information from the whole day. The calculation is described in section 2.2.2. To present the data in a more intuitive manner, the annualized VIX is divided by $\sqrt{252}$ (as for example in Blair et al. (2001) and Whaley (2008)).
The SPX together with the VIX is illustrated in Figure 1 and the realized volatility is added in Figure 2 in the appendix. The summary statistics can be found in Table 2 and using logarithm in Table 3, equally in the appendix. The summary statistics and figures show, that the VIX is for every time period slightly higher than the realized variance. This is consistent with the findings previous research, e.g. Jiang and Yisong (2005). However, the graphics show, that during crisis periods, the realized volatility exceeds the VIX. For example, whereas both the realized volatility and the VIX are particularly high during the crisis around 2008, the realized variance is peaking even higher. To account for this period, a dummy for the crisis 2008 - 2012 was included in the model. Moreover the summary statistics show, that the skewness and kurtosis of the logarithmic specification is closer to the one of the normal distribution. Consequently, a regression based on the log volatilities is also calculated and should be statistically better specified than the one based on simple volatility. The correlation matrix for the realized volatilities and model-free implied volatility can be found in Table 4 in the appendix. Overall, the realized volatility is highly correlated with both the past realized volatility and the past VIX values. For comparison, also the weekly and monthly VIX is included, calculated in the same way as the weekly and monthly realized volatility. Whereas for the one-day lag the correlation of the realized volatility with the

VIX is higher, for the lagged weekly and monthly averages the correlation with the historic volatilities is higher.

3.2 Methodology: Linear Regression and HAR-RV model

Consistent with prior research, for example Jiang and Yisong (2005), Canina and Figlewski (1993) or Christensen and Prabhala (1998), both univariate and encompassing regression analysis is used to analyse the information content of the volatility measures. Additionally the approach from the HAR-RV model, described by Corsi (2009) is integrated and not only one day lagged realized volatility is used as an explanatory variable in the regression, but also weekly and monthly realized volatility.

In the univariate regression, the realized volatility is regressed once only on the historic data, and once only on the VIX. For comparison, realized volatility is regressed on both historic data and the VIX in the encompassing regression analysis. Thus the encompassing regression analysis gives information about the relative importance of the volatility measures, and whether the VIX subsumes the information from the historic volatility (Jiang & Yisong, 2005). For each explanatory variable, the value of the current day t is used, whereas for the explained variable, the one-day ahead value $t+1d$ is used. Like this, all the information available on day t is used to evaluate the one-day ahead realized volatility. The three regressions are then given by

$$\begin{aligned} RV_{t+1d} = c &+ \beta_t^{RV,d} RV_t^{(d)} + \beta_t^{RV,w} RV_t^{(w)} + \beta_t^{RV,m} RV_t^{(m)} \\ &+ \beta^{crisis} crisis \end{aligned} \tag{Reg1a}$$

$$RV_{t+1d} = c + \beta_t^{VIX} VIX_t + \beta^{crisis} crisis \tag{Reg2a}$$

$$\begin{aligned} RV_{t+1d} = c &+ \beta_t^{RV,d} RV_t^{(d)} + \beta_t^{RV,w} RV_t^{(w)} + \beta_t^{RV,m} RV_t^{(m)} \\ &+ \beta_t^{VIX} VIX_t + \beta^{crisis} crisis \end{aligned} \tag{Reg3a}$$

with RV_{t+1d} the realized volatility one-day-ahead, $RV_t^{(d)}$ the daily realized volatility, $RV_t^{(w)}$ the weekly realized volatility, $RV_t^{(m)}$ the monthly realized volatility, VIX the VIX closing value and $crisis$ the dummy variable, indicating one in the time of the financial crisis (2008-2012) and zero otherwise. The same regressions are specified with the logarithm for each variable,

$$\begin{aligned} \ln(RV_{t+1d}) = c + \beta_t^{RV,d} \ln(RV_t^{(d)}) + \beta_t^{RV,w} \ln(RV_t^{(w)}) \\ + \beta_t^{RV,m} \ln(RV_t^{(m)}) + \beta^{crisis} crisis \end{aligned} \tag{Reg1b}$$

$$ln(RV_{t+1d}) = c + \beta_t^{VIX} ln(VIX_t) + \beta^{crisis} crisis \tag{Reg2b}$$

$$\begin{aligned} ln(RV_{t+1d}) = c + \beta_t^{RV,d} ln(RV_t^{(d)}) + \beta_t^{RV,w} ln(RV_t^{(w)}) + \beta_t^{RV,m} ln(RV_t^{(m)}) \\ + \beta_t^{VIX} ln(VIX_t) + \beta^{crisis} crisis. \end{aligned} \tag{Reg3b}$$

In all regression specifications the Newey-West covariance correction is used, to account for the possible presence of serial correlation in the data.

3.3 Research Questions

Model-free implied volatility, contrary to historic volatility, uses the forward-looking nature of options, does not suffer from misspecification problem and aggregates information across all strike prices. It is therefore expected, that it is informationally more efficient than historic volatility. BS implied volatility is not included, as previous research found that model-free implied volatility outperforms BS implied volatility (e.g. Bakanova (2010), Jiang and Yisong (2005)).

The regressions will be examined with the following research questions:

Q1: Does the VIX contain statistically significant information about one-day ahead realized volatility?

Q2: Does the VIX have more explanatory value than the historic volatilities in estimating one-day ahead realized volatility?

Q3: Does VIX add explanatory value to the historic volatilities in estimating one-day ahead realized volatility?

Q4: Does the VIX incorporate all information regarding one-day ahead realized volatility, hence the historic volatilities contain no information beyond what is already included in the VIX?

The research questions are approached in the following way:

Concerning **Q1**, if the VIX contains information about future volatility, we would expect the slope estimate of the VIX in Reg2a and Reg2b to be significantly different from zero. This can be formalized as a testable hypothesis, with $H_0 : \beta^{VIX} = 0$, using t-tests. For **Q2**, adjusted R^2 should be compared in the univariate regressions (Reg1a to Reg2a and Reg1b to Reg2b). If the VIX can explain more variation of the one-day-ahead realized volatility than the historic volatilities, the R^2 and adjusted R^2 should be larger in the second regression[4]. For **Q3**, adjusted R^2 should be compared in the univariate regression with the VIX and the regression containing all variables (Reg2a to Reg3a and Reg2b to Reg3b). If the VIX adds explanatory value additional to the historic volatilities, the (adjusted) R^2 should increase in the third specification. Finally, for **Q4**, if the VIX subsumes the information content contained in the historic volatilities, we would expect the slope estimates for the historic volatilities to become not significantly different from zero, when the VIX is added, whereas the VIX should still be significant (Reg3a and Reg3b). This can be formalized as a testable hypothesis with the $H_0 : \beta^d = \beta^w = \beta^m = 0 \textit{ and } \beta^{VIX} = 1$, using F-tests.

[4]For this paper the adjusted R^2 is preferred to the normal R^2, as it adds a penalty for adding additional independent variables, which could be misleading here because the first specification has more explanatory variables than the second one.

3.4 Limitations

This section will briefly present some limitations of the methodology applied, which is supposed to allow the reader to better understand the results.

As described multiple times, volatility can not directly be observed, even though futures and options on the VIX are now being increasingly traded. Consequently, the realized volatility used in this paper is a measurement over an aggregated period of discrete time. The approach used for realized volatility is shown to be a consistent nonparametric measure of notional volatility by Andersen et al. (2002), nevertheless also the notional volatility is only the volatility of a fixed time interval, thus it can not be interpreted as an instantaneous volatility measure.

The VIX was taken as it is calculated by CBOE, as the aim of this paper is not to improve any model-free implied volatility calculation, but to examine its information content. Such an improvement of the VIX calculation is done by Jiang and Tian (2007), as they state that the VIX significantly underestimates true volatility.

Considering the measurement, the methodology used aims at examining the information content and is an in-sample forecast, thus it can not be used directly for predicting unknown realized volatility.

4 Results

This section presents the estimation results. The first subsection reports the results obtained with the regression analysis described in section 3.2 and the second subsection provides the results of the robustness checks conducted.

4.1 Regression Analysis Results

The regression results can be found in Table 1 for the logarithmic regression and in Table 5 in the appendix for the level regression. To make it easier to overview the results, only one regression table was included in this section. The logarithmic table is displayed, since as discussed in section 3.1 it should be better specified, which is confirmed by it having a lower AIC than the level specification.

Firstly, considering the overall results, it is striking that the slope coefficients are positive and significantly different from zero for all volatility measures in all three regressions. This implies that they all contain substantial information for future volatility. Thus, also the slope of the VIX in Reg2a and Reg2b is positive and significantly different from zero, and the $H_0 : \beta^{VIX} = 0$ can be strongly rejected for both the level and the logarithmic regression, indicating that the VIX has statistically significant information content for one-day ahead realized volatility.

Secondly, if the VIX has more explanatory power than the historic volatilities, the adjusted R^2 in the second regression with the VIX (Reg2a and Reg2b) should be larger than in the first regression with the historic volatilities (Reg1a and Reg1b). This is not true for both the level and the logarithmic specification. Even though the R^2s are both high and close, it is slightly larger in the first regression (0.726) than in the second (0.693).

Thirdly, if the VIX adds explanatory value to the historic volatilities, the adjusted R^2 in the regression with the VIX included additionally to the historic volatilities (Reg3a and Reg3b) should be larger than in the regression containing only the historic volatilities (Reg1a and Reg1b). This is true for both specifications, the full model is the one with the highest (adjusted) R^2 (and AIC), hence the VIX has statistically significant information content which is not contained in the three historic volatilities..

Finally, if the VIX subsumes all information contained in the historic volatilities, the historic volatilities should not be significantly different from zero in the regression containing all explanatory variables, whereas the VIX should be significant. Even though the estimates for all historic volatilities decrease when the VIX is added, only the monthly historic volatility turns insignificant in the logarithmic specification. This is confirmed by the results from the F-test, which can be found in the appendix in Table 8 and Table 9. The $H_0 : \beta^d = \beta^w = \beta^m = 0 \; and \; \beta^{VIX} = 1$ can be rejected at the 0.001 significance level for both specifications. Thus, it can not be said that the VIX subsumes the information contained in the historic volatilities.

The estimate for the crisis dummy is also significantly different from zero in every regression in both the level and the logarithmic specification, thus including the crisis was reasonable. This is in line with Jiang and Yisong (2005), who state that it is critical to account for periods with particular financial turmoil, such as the regime shift around the October 1987 crash.

Table 1: Logarithmic regression (whole sample)

	Dependent variable:		
	Realized Volatility		
	Reg1b	Reg2b	Reg3b
	(1)	(2)	(3)
Intercept	-0.043^{***}	-0.407^{***}	-0.187^{***}
	(0.007)	(0.017)	(0.014)
$ln(RV_t^{(d)})$	0.344^{***}		0.263^{***}
	(0.027)		(0.025)
$ln(RV_t^{(w)})$	0.395^{***}		0.284^{***}
	(0.035)		(0.035)
$ln(RV_t^{(m)})$	0.208^{***}		0.014
	(0.024)		(0.029)
$crisis$	0.020^{*}	-0.224^{***}	-0.100^{***}
	(0.012)	(0.034)	(0.016)
$ln(VIX_t)$		1.472^{***}	0.648^{***}
		(0.048)	(0.045)
AIC	1874.2	2363.6	1551.6
Observations	4,434	4,434	4,434
R^2	0.726	0.694	0.745
Adjusted R^2	0.726	0.693	0.745
Residual Std. Error	0.299 (df = 4429)	0.316 (df = 4431)	0.288 (df = 4428)
Note:			*p<0.1; **p<0.05; ***p<0.01

The numbers in the brackets are the standard errors of the parameters computed with Newey-West covariance correction, which are robust to autocorrelated and heteroscedastic error terms, see Newey and West (1987).

4.2 Robustness Checks

Early studies (for example Canina and Figlewski (1993)) testing the information content of (model-free) implied volatility often used overlapping samples, meaning that the same option is used in several implied-volatility calculations. However, Christensen, Hansen and Prabhala (2002) showed that the use of overlapping samples creates a telescopic overlap problem under which standard statistical inferences (including t-statistic and R^2) are no longer valid.

Therefore the same regression analysis were conducted using non-overlapping samples. Jiang and Yisong (2005) create monthly non-overlapping samples, using the first Wednesday of every month, since they calculate the implied volatility over a horizon of one month. The VIX however is calculated slightly differently. It contains near- and next-term options options between 23 and 37 days to maturity (which is always a Friday), and every week the options roll over to new maturities. This shall be illustrated taking the example from Chicago Board of Options Exchange (2009). If for example the second Tuesday in October is taken, the near-term option expires in 24 days, and the next-term option in 31 days. One day later, the option that expires now in 30 days is the near-term option, and another option expiring in 37 days is the next-term option. This next-term option will, one week later, roll over to a near-term option and, one more week later, drop out of the calculation. Thus, an option can be included in the calculation for up to two weeks.

To avoid that an option can appear in two volatilities, after the volatilities and their lags are calculated, only one value every two weeks is used. As in Jiang and Yisong (2005), the values of Wednesday are used, having only few holiday days also in the sample of this paper, for each second week.

The estimation results for the sample using non-overlapping data are summarized in Table 6 and Table 7 in the appendix, the respective F-tests are in 10 and 11, also in the appendix.

The regression results using non-overlapping samples confirm all the results described for the regressions using the full sample. Reg2a and Reg2b show that the VIX contains significant information about future volatility. The increase of the adjusted R^2 from the second to the third regression, and the significance of the VIX in the full regression indicate that the VIX has explanatory value beyond that contained in the historic volatilities. On the other side, it can not be said that the VIX has a more explanatory power than the historic volatilities or that the VIX subsumes all the information contained in the historic volatilities.

Concerning the quality and predictive power of the model, also for this sample the AIC is lower in the logarithmic specification and the adjusted R^2, even though high in both specifications (0.796 and 0.770), is higher in the logarithmic regression. Within the specifications, it is also always the full model with the highest information content for one-day ahead realized volatility.

Nevertheless, there are two differences. Firstly, in the full-sample regression all estimates for the historic volatilities decreased when the VIX was added, and the monthly historic volatility turned insignificant only in the logarithmic specification. In the non-overlapping sample regression all estimates of historic volatilities decrease, too, but the monthly historic volatility turned insignificant in both the level and the logarithmic specification. Thus, apparently the VIX subsumes this information. Secondly, in the non-overlapping sample regressions both the intercept and the estimate for the crisis dummy are insignificant in the first regression, containing only the historic volatilities.

5 Conclusion

This paper analyses the information content for one-day ahead realized volatility of the VIX, as an example for model-free implied volatility, and compares it to historic volatilities, using daily, weekly and monthly historic volatility. After introducing different methods for volatility measurement and modelling, OLS regression is used to evaluate the information content. Apart from extending the regression approach from previous research, the paper follows the methodology corrections proposed to improve the statistical validity of the results, namely using long time series, high-frequency returns, non-overlapping samples and the inclusion of financial crisis periods.

The results show, that the VIX contains significant information content for future realized volatility, also beyond the information included in the historic volatilities. These findings are consistent with previous research and robust to serial correlation and alternative sampling methods.

Moreover, this paper shows, that not only the daily, but also the weekly average of the historic volatility contains significant information content for future volatility and it is useful to include them in the model. Comparing the models, the logarithmic specification always performed better than the level specification, which was to be expected seeing the descriptive statistics of the data.

However, it can not be concluded, that the model-free implied volatility subsumes all the information contained in the historic volatility, which partly contradicts previous research. The discrepancy could be due to several reasons. This paper used the approach from the HAR-RV model, but as daily volatility was significant in every regression that is not likely to be the reason. Alternatively, an explanation could be the different and longer time period considered, including and accounting for a financial crisis. Moreover, Jiang and Yisong (2005) could show only in the regression using the variances, that the model-free implied volatility subsumed the information contained in historic volatility. With volatilities, the model-free implied volatility was insignificant in the model containing all explanatory variables, and the R^2 did not increase either.

To conclude, the results show, that model-free implied volatility and the VIX contain significant information content for future realized volatility, but they do not fully subsume the information content from historic volatilities.

An interesting outlook for future research could be to further test the informational efficiency of model-free implied volatility using not index, but single stock data, as Taylor et al. (2010) did, and to further investigate the reason for the slight discrepancies found in current research.

References

Andersen, T. G. & Bollerslev, T. (1998). Answering the skeptics: Yes, standard volatility models do provide accurate forecasts. *International Economic Review*, *39*, 885–905.

Andersen, T. G., Bollerslev, T. & Diebold, F. X. (2002). *Parametric and nonparametric volatility measurement* (tech. rep. No. 279). NBER Technical Working Paper.

Andersen, T. G., Bollerslev, T., Diebold, F. X. & Labys, P. (2003). Modeling and forecasting realized volatility. *Econometrica*, *71*(2), 579–625. accessed 26-09-2018. Retrieved from https://www.jstor.org/stable/3082068

Bakanova, A. (2010). The information content of implied volatility in the crude oil market. In *Proceedings of the brunel conference* (pp. 1–19). Citeseer.

Black, F. & Scholes, M. (1973). The pricing of options and corporate liabilities. *Journal of Political Economy*, *81*(3), 637–654.

Blair, B. J., Poon, S.-H. & Taylor, S. J. (2001). Modelling S&P 100 volatility: The information content of stock returns. *Journal of Banking & Finance*, *25*(9), 1665–1679.

Britten-Jones, M. & Neuberger, A. (2000). Option prices, implied price processes, and stochastic volatility. *The Journal of Finance*, *55*(2), 839–866.

Canina, L. & Figlewski, S. (1993). The informational content of implied volatility. *The Review of Financial Studies*, *6*(3), 659–681.

Chicago Board of Options Exchange. (2009). *White paper Cboe volatility index*. Cboe Exchange.

Chicago Board of Options Exchange. (2018a). Historic daily prices. Data retrieved from http://www.cboe.com/products/stock-index-options-spx-rut-msci-ftse/s-p-500-index-options/s-p-500-index/spx-historical-data [accessed November 8, 2018].

Chicago Board of Options Exchange. (2018b). VIX data for 1990 - 2003. Data retrieved from http://www.cboe.com/products/vix-index-volatility/vix-options-and-futures/vix-index/vix-historical-data [accessed October 31, 2018].

Chicago Board of Options Exchange. (2018c). VIX data for 2004 to present. Data retrieved from http://www.cboe.com/products/vix-index-volatility/vix-options-and-futures/vix-index/vix-historical-data [accessed October 31, 2018].

Christensen, B. J., Hansen, C. S. & Prabhala, N. R. (2002). *The telescoping overlap problem in options data*.

Christensen, B. J. & Prabhala, N. R. (1998). The relation between implied and realized volatility1. *Journal of Financial Economics*, *50*(2), 125–150.

Corsi, F. (2009). A simple approximate long-memory model of realized volatility. *Journal of Financial Econometrics*, *7*(2), 174–196. doi:10.1093/jjfinec/nbp001

Derman, E. & Kani, I. (1994). Riding on a smile. *Risk*, *7*(2), 32–39.

Dupire, B. (1994). Pricing with a smile. *Risk*, *7*(1), 18–20.

Dupire, B. (1997). Pricing and hedging with smiles. *Mathematics of Derivative Securities*, *1*(1), 103–111.

Fleming, J., Ostdiek, B. & Whaley, R. E. (1995). Predicting stock market volatility: A new measure. *The Journal of Futures Markets*, *15*(3), 265–302.

Jiang, G. J. & Tian, Y. S. (2007). Extracting model-free volatility from option prices: An examination of the vix index. *Journal of Derivatives*, *14*(3).

Jiang, G. J. & Yisong, T. S. (2005). Model-free implied volatility and its information content. *The Review of Financial Studies*, *18*(4). doi:10.1093/rfs/hhi027

Malkiel, B. G. & Fama, E. F. (1970). Efficient capital markets: A review of theory and empirical work. *The Journal of Finance*, *25*(2), 383–417.

Mandelbrot, B. B. (1963). The variation of certain speculative prices. *The Journal of Business*, *36*, 394–419.

Müller, U. A., Dacorogna, M. M., Davé, R. D., Pictet, O. V., Olsen, R. B. & Ward, J. R. (1993). Fractals and intrinsic time: A challenge to econometricians. Applied Econometrics Asociation.

Newey, W. K. & West, K. D. (1987). Hypothesis testing with efficient method of moments estimation. *International Economic Review*, 777–787.

Oxford-Man Institute's realized library V0.2. (2009). Data retrieved from https://realized.oxford-man.ox.ac.uk/data [accessed October 13,2018].

Poon, S.-h. & Granger, C. W. (2003). Forecasting volatility in financial markets: A review. *Journal of Economic Literature*, *41*(2), 478–539.

Rubinstein, M. (1994). Implied binomial trees. *The Journal of Finance*, *49*(3), 771–818.

Taylor, S. J., Yadav, P. K. & Zhang, Y. (2010). The information content of implied volatilities and model-free volatility expectations: Evidence from options written on individual stocks. *Journal of Banking & Finance*, *34*(4), 871–881.

Tsay, R. S. (2005). *Analysis of financial time series*. John Wiley & Sons.

Whaley, R. E. (2008). *Understanding vix*. The Owen Graduate School of Management, Vanderbilt University. [accessed October 14, 2018]. Retrieved from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1296743

Appendix

Figures

Figure 1: S&P 500 and VIX

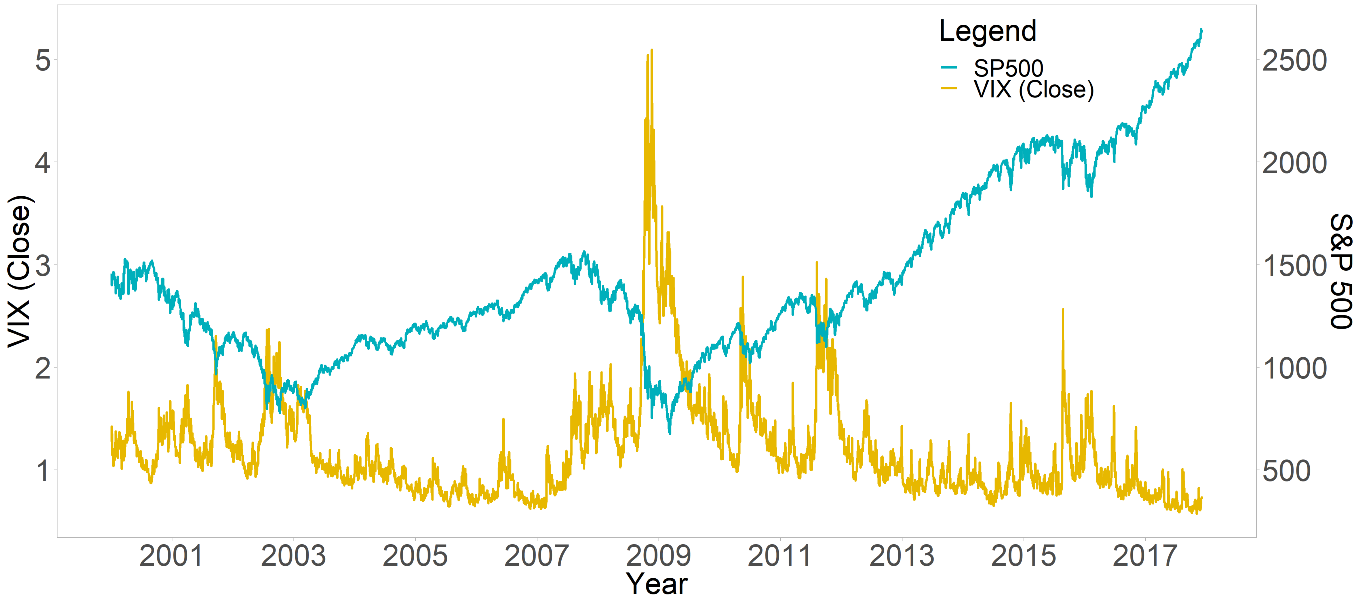
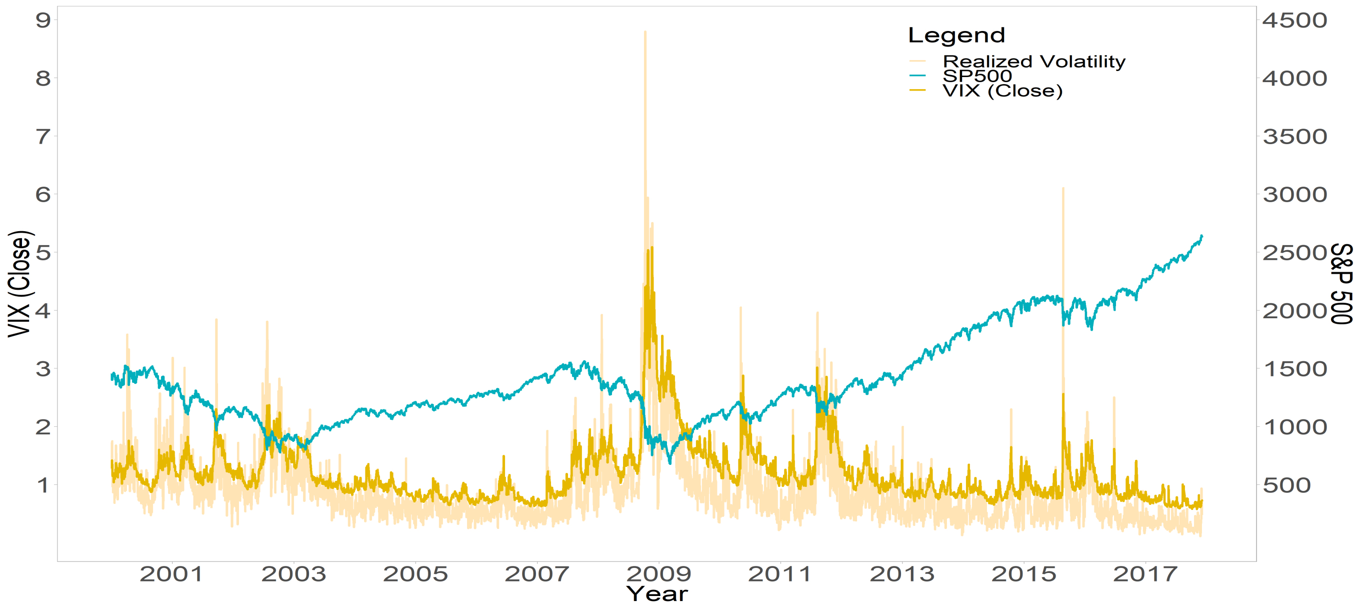


Figure 2: S&P 500, RV and VIX



Data Summary Statistics

Table 2: Summary statistics: Variables

Statistic	N	Min	Max	Mean	St. Dev.	Skewness	Kurtosis
All time periods							
RV	4,434	0.110	8.802	0.874	0.615	3.127	17.751
VIX	4,434	0.576	5.094	1.196	0.525	2.552	9.787
Weekly RV	4,434	0.183	5.586	0.874	0.556	2.757	12.451
Monthly RV	4,434	0.223	4.375	0.876	0.512	2.534	10.051
Weekly VIX	4,434	0.596	4.593	1.197	0.519	2.510	9.235
Monthly VIX	4,434	0.618	4.126	1.198	0.504	2.491	8.853
During crisis							
RV	1,252	0.213	8.802	1.171	0.834	2.713	11.399
VIX	1,259	0.847	5.094	1.623	0.695	1.941	4.232
Weekly RV	1,259	0.257	5.586	1.169	0.753	2.409	7.417
Monthly RV	1,259	0.423	4.375	1.172	0.689	2.178	5.448
Weekly VIX	1,259	0.885	4.593	1.623	0.684	1.885	3.753
Monthly VIX	1,259	0.946	4.126	1.625	0.662	1.837	3.311
Outside of crisis							
RV	3,229	0.110	6.109	0.762	0.454	2.252	10.798
VIX	3,251	0.576	2.566	1.034	0.307	1.206	1.365
Weekly RV	3,246	0.183	3.165	0.765	0.400	1.564	3.441
Monthly RV	3,229	0.223	2.367	0.764	0.363	1.308	1.687
Weekly VIX	3,246	0.596	2.188	1.034	0.300	1.157	1.047
Monthly VIX	3,229	0.618	2.014	1.033	0.285	1.1	0.78

Table 3: Summary statistics: Logarithm variables

Statistic	N	Min	Max	Mean	St. Dev.	Skewness	Kurtosis
			All time periods				
RV	1,236	−9.664	0.777	−1.469	1.262	-1.369	3.185
VIX	2,528	−8.005	0.487	−1.562	1.190	-1.294	2.586
Weekly RV	4,434	0.183	5.586	0.874	0.556	2.757	12.451
Monthly RV	4,434	−1.501	1.476	−0.259	0.483	0.487	0.357
Weekly VIX	4,434	−0.518	1.525	0.111	0.350	0.933	1.117
Monthly VIX	4,434	−0.481	1.417	0.115	0.340	0.963	1.227
			During crisis				
RV	1,252	0.213	8.802	1.171	0.834	2.713	11.399
VIX	1,259	0.847	5.094	1.623	0.695	1.941	4.232
Weekly RV	1,259	0.257	5.586	1.169	0.753	2.409	7.417
Monthly RV	1,259	0.423	4.375	1.172	0.689	2.178	5.448
Weekly VIX	1,259	0.885	4.593	1.623	0.684	1.885	3.753
Monthly VIX	1,259	0.946	4.126	1.625	0.662	1.837	3.311
			Outside of crisis				
RV	3,229	0.110	6.109	0.762	0.454	2.252	10.798
VIX	3,251	0.576	2.566	1.034	0.307	1.206	1.365
Weekly RV	3,246	0.183	3.165	0.765	0.400	1.564	3.441
Monthly RV	3,229	0.223	2.367	0.764	0.363	1.308	1.687
Weekly VIX	3,246	0.596	2.188	1.034	0.300	1.157	1.047
Monthly VIX	3,229	0.618	2.014	1.033	0.285	1.100	0.780

Table 4: Correlation table

	RV	VIX	Daily RV	Weekly RV	Monthly RV	D. VIX	W.VIX	M. VIX
RV	1	0.837	0.806	0.823	0.769	0.819	0.781	0.698
VIX	0.837	1	0.822	0.883	0.903	0.981	0.969	0.926
Daily RV	0.806	0.822	1	0.890	0.794	0.837	0.803	0.710
Weekly RV	0.823	0.883	0.890	1	0.903	0.896	0.906	0.809
Monthly RV	0.769	0.903	0.794	0.903	1	0.910	0.932	0.931
Daily VIX	0.819	0.981	0.837	0.896	0.910	1	0.982	0.935
Weekly VIX	0.781	0.969	0.803	0.906	0.932	0.982	1	0.961
Monthly VIX	0.698	0.926	0.710	0.809	0.931	0.935	0.961	1

Regression Results with Robustness Checks

Table 5: Level regression (whole sample)

	Dependent variable:		
	Realized Volatility		
	Reg1a	Reg2a	Reg3a
	(1)	(2)	(3)
Intercept	0.045***	−0.324***	−0.169***
	(0.015)	(0.059)	(0.034)
$RV_t^{(d)}$	0.362***		0.256***
	(0.038)		(0.040)
$RV_t^{(w)}$	0.391***		0.286***
	(0.056)		(0.064)
$RV_t^{(m)}$	0.188***		−0.106**
	(0.036)		(0.050)
$crisis$	0.025*	−0.214***	−0.112***
	(0.013)	(0.035)	(0.021)
VIX_t		1.052***	0.579***
		(0.059)	(0.064)
AIC	2817.4	3104.2	2446
Observations	4,434	4,434	4,434
R^2	0.708	0.689	0.732
Adjusted R^2	0.708	0.688	0.732
Residual Std. Error	0.332 (df = 4429)	0.343 (df = 4431)	0.319 (df = 4428)
Note:			*p<0.1; **p<0.05; ***p<0.01

The numbers in the brackets are the standard errors of the parameters computed with Newey-West covariance correction, which are robust to autocorrelated and heteroscedastic error terms, see Newey and West (1987).

Table 6: Level regression (non-overlapping sample)

	Dependent variable:		
	Realized Volatility		
	Reg1a	Reg2a	Reg3a
	(1)	(2)	(3)
Intercept	0.046	-0.342^{***}	-0.134^{**}
	(0.031)	(0.091)	(0.052)
$RV_t^{(d)}$	0.408^{***}		0.330^{***}
	(0.109)		(0.116)
$RV_t^{(w)}$	0.501^{***}		0.394^{***}
	(0.126)		(0.128)
$RV_t^{(m)}$	0.072		-0.132
	(0.079)		(0.083)
$crisis$	-0.022	-0.257^{***}	-0.131^{***}
	(0.029)	(0.058)	(0.035)
VIX_t		1.092^{***}	0.460^{***}
		(0.094)	(0.089)
AIC	192.5	281.8	166
Observations	456	456	456
R^2	0.757	0.701	0.771
Adjusted R^2	0.754	0.700	0.769
Residual Std. Error	0.297 (df = 451)	0.328 (df = 453)	0.288 (df = 450)
Note:			*p<0.1; **p<0.05; ***p<0.01

The numbers in the brackets are the standard errors of the parameters computed with Newey-West covariance correction, which are robust to autocorrelated and heteroscedastic error terms, see Newey and West (1987).

Table 7: Logarithmimc regression (non-overlapping sample)

	Dependent variable:		
	Realized Volatility		
	Reg1b	Reg2b	Reg3b
	(1)	(2)	(3)
Intercept	-0.001	-0.365^{***}	-0.156^{***}
	(0.018)	(0.022)	(0.029)
$ln(RV_t^{(d)})$	0.346^{***}		0.262^{***}
	(0.064)		(0.062)
$ln(RV_t^{(w)})$	0.408^{***}		0.289^{***}
	(0.078)		(0.076)
$ln(RV_t^{(m)})$	0.171^{***}		-0.034
	(0.054)		(0.064)
crisis	-0.017	-0.266^{***}	-0.148^{***}
	(0.027)	(0.046)	(0.034)
$ln(VIX_t)$		1.478^{***}	0.700^{***}
		(0.060)	(0.104)
AIC	126.4	173.4	81.8
Observations	456	456	456
R^2	0.748	0.718	0.773
Adjusted R^2	0.746	0.717	0.770
Residual Std. Error	0.276 (df = 451)	0.291 (df = 453)	0.262 (df = 450)
Note:			*p<0.1; **p<0.05; ***p<0.01

The numbers in the brackets are the standard errors of the parameters computed with Newey-West covariance correction, which are robust to autocorrelated and heteroscedastic error terms, see Newey and West (1987).

F-test Results with Robustness Checks

Table 8: F-test Reg3a

	Res.Df	RSS	Df	Sum of Sq	F	Pr(>F)
1	4432	524.31				
2	4428	449.29	4	75.03	184.86	0.0000

Table 9: F-test Reg3b

	Res.Df	RSS	Df	Sum of Sq	F	Pr(>F)
1	4432	529.46				
2	4428	367.22	4	162.24	489.09	0.0000

Table 10: F-test Reg3a non-overlapping sample

	Res.Df	RSS	Df	Sum of Sq	F	Pr(>F)
1	454	49.39				
2	450	37.26	4	12.13	36.62	0.0000

Table 11: F-test Reg3b non-overlapping sample

	Res.Df	RSS	Df	Sum of Sq	F	Pr(>F)
1	454	47.32				
2	450	30.98	4	16.34	59.33	0.0000

Declaration of Authorship

I hereby declare that the paper

“The Information Content of VIX Volatility"

is my own unaided work. All direct or indirect sources used are acknowledged as references. I am aware that the paper in digital form can be examined for the use of unauthorized aid and in order to determine whether the thesis as a whole or parts incorporated in it may be deemed as plagiarism. This paper was not previously presented to another examination board and has not been published.

Place, Date

author’s signature