Zeppelin Universität

Chair of Empirical Finance and Econometrics

Measuring the information content of VIX volatility in comparison to historic volatility

Humboldt Research Project Project

in

Module 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

Glossary

BS Black-and-Scholes-Merton Model. 3, 4

Abstract

This paper investigates the information content of model-free implied volatility for daily realized volatility of the S&P 500, using the VIX from the Chicago Board of Options Exchange. In contrast to the Black-and-Scholes implied volatility, the VIX is not based on any specific option pricing model and therefore does not suffer from the joint hypothesis problem and provides a direct test for market efficiency. Using the approach from an HAR-RV model as described by Corsi (2009) to account for the multi-scaling observed in financial data, the paper conducts an encompassing regression analysis to address the research question. The results show, that the VIX subsumes all the information contained in historic volatility. This results are be consistent and robust to serial correlation.

1 Introduction: The Importance of Volatility Measurement

1.1 Why Volatility matters: Volatility as Key Input to Option Pricing Models and Risk Measures

Distributional characteristics of asset returns are of high interest for the financial sector. They are for example key input to the pricing of financial instruments like derivatives, or to risk measures, such as the Value-at-Risk. Moreover they give information on the risk-return trade-off, which is a central question in portfolio allocation and managerial decision making. Of particular interest is the asset's return volatility, being the most dominant time-varying distribution characteristic. (Andersen et al. 2003).

Risk is important because.. . The concept of risk is closely linked to that of volatility, as the second moment characterstic of asset return distributions. Many pricing models measure risk through volatility, which thus influences the expected return (Harvey and Whaley 1992). Moreover many risk-measure, such as the Value-at-Risk are very closely related to volatility. (Alternative)

Volatility "seeks to capture the strength of the (unexpected) return variation over a given period of time" (**andersen2001**)

As volatility is not directly observable, it has to be estimated. During the last years, considerable research has been devoted to the question, how volatility can be measured or estimated. Two prominent categories of approaches are on the one hand time series models using historic volatility, and on the other hand implied models using option price data[1]. In the recent years, Black-and-Scholes implied volatility measurement gained popularity, this approach uses the forward-looking nature of option prices. Options are contracts, giving the holder the right to either buy (call option), or sell (put option), an underlying

[1]There are, of course, various other methods, such as nonparametric methods or neural networks based models (Jiang, George J & Tian 2003), however they shall not be discussed here

asset, at a specified date in the future for a certain price (John et al. 2006). Assuming rational agents/expectation, the market uses all available information to form it's expectation about future price movements and thus about volatility. Assuming furthermore that the market is efficient (meaning as Eugene Fama defined it, that prices reflect all available information), the market's estimate of future volatility is the best possible forecast possible, given the current information (Bent Jesper Christensen and Hansen 2002). Due to this forward looking component of option contracts, option prices indirectly contain the market participants' expectations of the underlying asset's future movements. A widely used model to price this option contracts is the Black-and-Scholes-Merton model, which uses the option's volatility as an input factor. By 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 volatiltiy of the underlying asset" (Jiang, George J & Tian 2003, p.1305).

Early studies found implied volatility to be a biased forecast of realized volatility, not containing significantly more information than historic volatility. More recent studies however presented evidence that there is important information contained in option prices, that adds to the efficiency of volatility forecasting when implied volatility is included (ibid.). A reason for this discrepancy in results could be that early studies did not consider several data and methodological problems, such as long enough time series, a possible regime shift around the crash in 1987 and the use of non-overlapping samples (ibid.). Bent J Christensen and Prabhala for example took this into account and found that implied volatility outperforms historic volatility. All in all Jiang, George J & Tian summarize, that collectively "these studies present evidence that implied volatility a more efficient forecast for future volatility than historic volatility" (p.1306).

1.2 Weaknesses of Existing Models: VIX Introduced by CBOE

Even though BS implied volatility is found to be the overall more efficient forecast of realized volatility compared to historic volatility (Jiang, George J & Tian 2003), the BS implied volatility has some specification problems. Firstly, BS implied volatility focuses on at-the-money options. The advantage is, that at-the-money options are the once most actively traded and thus the most liquid ones. However this focus fails to include information contained in other options. Moreover, volatility estimation with the BS model, includes the same assumptions as are made in the BS model itself. Thus tests based on the BS equation are actually joined tests of market efficiency (as market efficiency is assumed to use option prices for volatility estimation, as mentioned above) and the BS model, and therefore suffer from a model misspecification error (ibid.).

That is why during the last years, implied volatility indexes which are not based on a pricing assumption gained popularity. One of these model-free implied volatility indexes is the VIX from CBOE.

- power of volatility models lies in out-of-sample forecasting power
- so far BS implied volatility models had the best out of sampling forecasting power, but they have several problems (most importantly joint hypothesis problem)

2 Selected volatility concepts and models of volatility measurement

This section presents first some stylized facts of financial data, and gives an introduction to the different ways to estimate volatility. By pointing out the advantages and disadvantages of the concepts and models and their fit to the stylized facts, the HAR-RV approach shall be motivated.

2.1 The Return Process and Stylized Facts of Financial Data

As mentioned in the introduction, the challenge when measuring volatility is, that stock return volatility is not directly observable (Tsay 2005). This problem evolves from the fact that we can only observe one realization of the underlying data generating process, and even though stocks are traded and thus have market prices which could be used for volatility measurement, there is no continuous data available and even for high-frequency data and extremely liquid markets microstructure effects and noise prevent getting close to a continuous sample path. It is thus only possible to estimate avearages of discrete volatility for a given period of time. (**andersen2001**).
There are however several approaches that should be introduced here. To start with, the definition of the simple gross return is

$$1 + R_t = \frac{P_t}{P_{t-1}} \tag{1}$$

In the continuous-time setting, continuously compounded returns are used, which are given by

$$r_t = ln(1 + R_t) = ln(\frac{P_t}{P_t} + \frac{P_{t-1} - P_t}{P_t}) = ln\frac{P_t}{P_{t-1}} = p_t - p_{t-1}\ with\ p_t = ln(P_t) \tag{2}$$

When observed over time, this asset returns show some distributional properties, often referred to as stylized facts of asset returns. Observed by many authors, only a few shall be mentioned here. Corsi for example mentions particularly the very strong persistence of autocorrelation of the square and absolute returns, which regularly poses challenges to econometric models. Moreover return probability density functions are often leptocurtic with fat tails. As the time scale increases, the return distribution slowly converge to the normal distribution, but before convergence, the return distribution has different shapes depending on the time scale. Financial data also show evidence of scaling as described firstly by Mandelbrot, which is connected to the idea that patterns appear in different times, or that the distribution for returns has similar functional forms for various choices of the time interval.

Moreover **andersen2001** mentions that first, even though raw returns have a leptocurtic distribution, the returns standardized by realized volatility are approximately Gaussian. Second, the distribution of realized volatility of returns itself is right skewed, the one of the logarithms of realized volatility however are also approximately Gaussian. Third, the long-run dynamics of realized logarithmic volatilities are well approximated by a fractionally-integrated long-memory process. Other authors who mention stylized facts: (Jiang, George J & Tian 2003). Moreover, even though volatility is not directly observable, it has some properties that are commonly seen. Here the characteristics as presented by Tsay should be introduced. These are firstly volatility clusters, loosely speaking meaning that high volatility tends to be followed by high volatility, and the same is true for low volatility periods. Moreover, volatility jumps are rare. Third, volatility varies within a fixed range and does not diverge to infinity, which statistically means that volatility is often (weakly) stationary. Fourth, there is a leverage effect in volatility, meaning that it reads differently to a big increase in price than to a drop.

2.2 Concepts and Models using Historic Volatility

2.2.1 Ex-post Volatility Measurement - Realized Volatility

By definition “volatility seeks to capture the strength of the (unexpected) return variation over a given period of time” (**andersen2001**). However, there are multiple concepts and definitions of asset volatility. According to

andersen2001 the concepts 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. The aim of the HAR-RV model introduced later is to model volatility ex-post and then maybe use this approach to measure the information content of implied volatility and maybe also to build a forecast model. Thus the model needs some measure of latent volatility process, so that the performance can be evaluated.

It can be shown, that under some assumptions, realized volatility as the square root of the sum of squared high frequency returns, can be used to approximate the quadratic variation process which is the variation in a continuous time setting. This approach mainly building on the work of **andersen2001** und [noch jemanden finden] shall only be briefly introduced here.

To begin with, it should be assumed that we have a continuous-time no-arbitrage setting. As return volatility aims to capture the strength of the unexpected return variation, one needs to define the component of a price change as opposed to an expected price movement. This requires the decomposition of the return process in an expected and an innovation component. **andersen2001** 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). As mentioned in the previous paragraph, volatility measures often focus on representing the average volatility over a discrete time, as a continuous record of price data is not available, and even for liquid markets it is distorted by microstructure effects. **andersen2001** show that in order to measure the average variance[2], one can refer to the quadratic variation process of this martingale component, as the quadratic variation process represents the (cumulative) realized sample path variability of the martingale over any fixed time interval. To be precise, they define *notional variance* as the increment to the quadratic variation for the return series, measured ex-post.

Assuming that the mean of the return process is zero, taking the expected value of the notional variance and extending this concept slightly, one gets the *realized variance*, defined over the $[t-h,t], 0 < h \leq t \leq T$ time interval as

$$v^2(t,h;n) = \sum_{i=1}^{n} r(t-h+(i/n)\times h, h/n)^2. \tag{3}$$

andersen2001 show not only that the realized variance is an *unbiased* estimator of ex-ante expected variance, or at least approximately unbiased when relaxing the zero mean assumption and taking a high sample frequency (their proposition 4). Moreover, **andersen2001** show that the realized variance is a *consistent* nonparametric measure of the notional variance for increasingly finely sampled returns over any fixed length interval (their proposition 5).

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

2.2.2 Volatility Model - HAR-RV Model

Having introduced the concept of notional volatility and it's approximation by realized volatility, we now turn to volatility modelling/measurement. To measure volatility, one can separate between parametric and non-parametric

[2]actually **andersen2001** term this the *realized volatility*, but this paper will use the terminology of Corsi (2009), who uses the term realized volatility for the square-root of the integrated variance.

methods. Whereas parametric methods try to measure the expected volatility making different assumptions about both the functional form and the variables in the information set available, non-parametric methods try to quantify notional volatility directly. The realized volatility is an example for a non-parametric methods. However, to estimate volatility ex-ante (and to potentially forecast later), this paper will refer to one type of the parametric methods based on the paper of Corsi (2009) termed the HAR-RV model.

In the last paragraph, volatility was measured as the average over discrete time. For the HAR-RV model this concept needs to be transferred to an instantaneous volatility measurement in continuous time. Moving to continuous time, the corresponding logarithmic price process is then given by Corsi (ibid.) as the following stochastic difference equation,

$$dp(t) = \mu(t)dt + \sigma(t)dW(t), \ 0 \leq t \leq T \tag{4}$$

where $p(t)$ is the logarithm of the instantaneous price, μ is a finite variation stochastic process, $W(s)$ standard Brownian motion and σ a stochastic process independent of $W(s)$. For the volatility the difference is, that now the notional variance equals the *integrated variance (IV)*, which is the integral of the stochastic process. For a period of one day, this is represented by

$$IV_t^{(d)} = \int_{t-1d}^{t} \sigma^2(w)dw. \tag{5}$$

Corsi (ibid.) then take the square-root of this integrated variance and term it *integrated volatility*. **andersen2001** show that also in this continuous-time setting the integrated variance equals the notional variance, and thus can also be approximated with the sum of squared returns. This leads to the following notion of realized volatility over the one day interval

$$RV_t^{(d)} = \sqrt{\sum_{j=0}^{M-1} r_t^2 - j \times \Delta} \tag{6}$$

with $\Delta = 1d/M$ being the sampling frequency and $r_t^2 - j \times \Delta$ defined as the continuously compounded Δ-frequency returns.

This notion of volatility is combined with the *Heterogeneous Market Hypothesis* by Müller et al. (1993). This approach grounds on the fractal approach of Mandelbrot's 1983 paper, where (time series) objects are analyzed on different time scales and the obtained results are compared. The starting point of 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. Thus for example a process explaining monthly price changes could not be applied to daily data and vice versa. This fractal approach led to the heterogeneous market hypothesis. This hypothesis states that the market gives rise to heterogeneous trading behaviours as 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 side the high-frequency dealers such as market makers, in the middle some medium term dealers and on the other side the low-frequency dealers, such as central banks or commercial organizations. As the market is driven by these components, it can be described as "fractal". Müller et al. (ibid.) base this theory of different time horizons on multiple observations: Firstly, 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, this in sum comes close to a hyperbolic decline. Secondly, 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). Other than to time scale, the heterogeneous market hypothesis can be applied to geographical location, degrees of risk aversion, institutional constraints or transactions costs. However, as in Corsi (2009) for this paper the time aspect should be relevant. Corsi (ibid.) refers to studies that measure volatility over different time horizons and observe that there is an asymmetric behaviour of volatility: “The volatility over longer time intervals has a stronger influence on volatiilty over shorter time intervals than converseley” (ibid., p.178). Thus a cascade pattern from low to high frequencies emerges. To formalize the model the latent partial volatility $\sigma_t^{(d)}$ is defined as the volatility generated by a certain market component (here daily). 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. Each of this volatility components corresponds to a market component, that forms the expectation for the volatility of the next period 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. This formalizes in a cascade model, written as

$$\sigma_{t+1d}^{(d)} = c + \beta^d RV_t^{(d)} + \beta^{(w)} RV_t^{(w)} + \beta^{(m)} RV_t^{(m)} + \tilde{w}_{t+1d}^{(d)}, \tag{7}$$

with $\tilde{w}_{t+1d}^{(d)}$ being the innovation term. Transferring this to a time series model, the ex-post approximation of latent volatility with realized volatility is used, which can be written as

$$\sigma_{t+1d}^{(d)} = RV_{t+1d}^{(d)} + w_{1+1d}^{(d)}. \tag{8}$$

where the $w_t^{(d)}$ now subsumes both the latent volatility measurement and estimation errors. Taking equations 7 and 8 together, we obtain the time series representation from the cascade model,

$$RV_{t+1d}^{(d)} = c + \beta^d RV_t^{(d)} + \beta^{(w)} RV_t^{(w)} + \beta^{(m)} RV_t^{(m)} + w_{t+1d}, \tag{9}$$

where $w_{t+1d} = \tilde{w}_{1+d}^{(d)} - w_{1+d}^{(d)}$. Using simulated data, Corsi (2009) shows, that the HAR-RV simulated 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, meaning that the fat tails get thinner as the aggregation level increases. Concerning the volatility memory, the simulated data is also able to reproduce the long memory of the empirical data. Moreover they use OLS and forecast to show [complete this section, saying that they also show that there is information content and forecasting ability].

2.3 Implied volatility

2.3.1 The General Idea of Implied Volatility

As mentioned, there are multiple ways to model and measure volatility. One approach is to use not only historic price data, but to augment the information set by incorporating options data and using their forward looking nature. This approach can be termed *implied volatility*, and can consist for example of a parametric volatility model for returns, accompanied by an asset pricing model and the option price information(**andersen2001**). The intuition behind this is, that option prices can be seen as reflecting market participants' expectations of the future movements of the underlying asset. Assuming that the marked is informationally efficient, as described by Malkiel and Fama (1970), and that the asset pricing model is correct, the implied volatility derived from this model should not only subsume all information contained in the models using historic volatility, but also be a more efficient forecast of future volatility (Jiang, George J & Tian 2003).

One popular example of implied volatility is the *Black-and-Scholes implied*

volatility, building on the Black-and-Scholes-Merton (BS) asset pricing model as presented by Black and Scholes (1973). One input for pricing options with the BS model is the volatility of the underlying asset. Having the derivative prices available on the market, it is possible to extract a value for the expected volatility, by inverting the theoretical asset pricing model. It is however important to note, that all of these procedures depend on the assumptions that are made in the asset pricing model (**andersen2001**).
Many studies examined the information content of BS implied volatility and it's forecasting ability for future volatility. Earlier studies find that implied volatility contains little additional information content compared to historical volatility and has no forecasting ability (Jiang, George J & Tian 2003). Canina and Figlewski (1993) for example found that implied volatility does not incorporate the information from historic volatility and constitute poor forecast of future volatility using S&P 500 options. In contrast, other studies support the evidence of implied volatility subsuming the information from historic volatility. These are for example Day and Lewis (1992), Lamoureux and Lastrapes (1993) or Jorion (1995). Day and Lewis (1992) examine information content from S&P 100 options between 1983 and 1989 and find that they have significant information content for realized volatility. With previous research leading ambiguous results, more 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 for example using longer time series extending the regime shift around the 1987 crash or adapting high-frequency asset returns to provide a more accurate view or use non-overlapping samples (Jiang, George J & Tian 2003). For example Bent J Christensen and Prabhala (1998) find that subsumes the information content of past volatility and outperforms it in forecasting future volatility. They use a longer time series, take into account the 1987 regime shift and use non-overlapping samples (Jiang, George J & Tian 2003). Moreover Bent Jesper Christensen, Hansen

and Prabhala (2001) found that under certain measurement errors such as overlapping samples statistical tests are no longer meaningful. Taking this measurement errors into account results in support of the hypothesis, that implied volatility subsumes all information contained in historic volatility. All in all, this insights provide a convincing argument why early studies found the forecast of implied volatility to be inefficient (Jiang, George J & Tian 2003). For a very comprehensive overview of the volatility research as was state of the art in 2003, please see **poon**.

However, even though the recent evidence is in support of BS implied volatility, there are several disadvantages of using the implied volatility with an asset pricing model. One disadvantage is, that the BS implied volatility relies mostly on information from at-the-money options, which are generally the most actively traded ones. This however fails to incorporate information contained in other options. Moreover and more importantly, testing for the BS implied volatility always means testing market efficiency and the BS model jointly. Assumptions of the BS model include constant volatility, no transaction costs or taxes, no dividend before option maturity, no arbitrage, continuous trading, constant risk-free interest rate, divisible securities and no short sell (S.-h. Poon and Granger 2003). As it is not possible in this setting to test for market efficiency or the pricing assumptions separately, this tests are subject to model misspecification errors (Jiang, George J & Tian 2003). Moreover most early paper's implied volatility ignores the early exercise opportunity and/or dividends (Blair, S.-H. Poon and Taylor 2001). This is why an implied volatility that does not rely on an asset pricing model was introduced.

2.3.2 VIX and Model-Free Implied Volatility

The idea of model free implied volatility was introduced by Britten-Jones and Neuberger (2000) who show that the risk-neutral realized volatiltiy can be derived from a set of options with matching expiration, thus extendending the approach of Derman and Kani (1994) Dupire et al. (1994), Dupire (1997) and Rubinstein (1994) on implied distributions (Jiang, George J & Tian 2003). The idea of model free implied volatility is, that insted of specifing a precess for the price of the underlying security and then to derive the option price as a function of this prices process parameters, a complete set of option prices is taken as given and then as much information as possible is extracted out of the underyling price process (Britten-Jones and Neuberger 2000). To be more specific, Britten-Jones and Neuberger (ibid.) show in an approach resembling a binominal 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. This leads them to their proposition 1, that the expectation of the squared return, conditional on stock price and time, is determined by the initial option prices. Moreover, as this proposition one only infers a one-period forecast conditional on a stock price level, they propose a forecast over any multiperiod interval without conditioning, by showing that the risk-neutral expected sum of squared returns between to dates is given from the set of options expiring on these to dates (their proposition 2). Thiis formula does not use any specific option pricing model to derive implied volatility and is only based on no-arbitrage conditions, therefore it solves the joint-hypothesis problem and test directly for market efficiency (Jiang, George J & Tian 2003).

Several papers used this result, examined whether the model free implied volatility subsumes the information from the historic and BS implied volatility, conducted forecasts using the model free implied volatility and extended the concept.

(write down papers that work with model free implied volatility, e.g. Jiang, George J & Tian extended the model so that is not derived under diffusion assumptions and generalized it to processes including random jumps.)

One of the first implemented implied volatility index was the VIX from the CBOE. It is computed each trading day on a real-time basis, and to facilitate comparison it was calculated back to January 1986. It's introduction in 1993 had the intention to provide 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 S&P 100 options, as at this time the S&P 100 options where the most actively traded in the U.S., which is critical to the usefullness of the VIX or any other implied volatiltiy index. During the years the index option market changed and in 2003 the VIX adopted to this change. First, the S&p 500 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 the S&P 500 options. Secondly, option index trading behavior changed. Whereas in the 1990s both call and put index options where equally important, over the years out-of-the money and at-the-money put gained popularity as they were bought by portfolio insurers. Thus the VIX also started to include out-of-the money options in it's calculation, bringing another advantage that was not included in other early implied volatiltiy indexes as mentioned above, which lacked this information conten (Whaley 2008). Finally, the VIX calculation changed, adopting to the model-free implied volatility approach, which was in 2003 widely used by financial theorists, risk managers, and volatility traders alike (**Cboe2009**). Sometimes the VIX is called *investor fear gauge*, however it is important to notice that it measures, not causes market volatility. It is true, that using regression analysis Whaley (ibid.) found, that the rate of change in the VIX and S&P 500 is asymmetric, with the VIX reacting higher to a drop in the S&P 500 than its rise, which could be interpreted as a higher fear in the downside then excitement in an up-move. Nevertheless, this correlation mus not express causality (ibid.). 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, S.-H. Poon and Taylor 2001). The formula for calculation is

$$VIX = \sqrt{\frac{2}{T}\sum_i \frac{\Delta K_i}{K_i^2} e^{RT} Q(K_i) - \frac{1}{T}(\frac{F}{K_0} - 1)^2} \times 100 \qquad (10)$$

with T being 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 and $Q(K_i)$ the midpoint of bid-ask spread for each option with strike K_i. For the calculation first the options used for calculation are selected. The VIX always uses near- and next-term put and call options, with more than 23 and less than 37 days to expiration. The VIX will always reflect an interpolation of the volatility between these two option maturities, using the midpoint bid-ask price as the transaction price are subject to bid-ask bounde (S.-h. Poon and Granger 2003). Once a week the options to calculate the VIX roll over to new maturities. Then the volatility for both the near- and next-term options is calculated. Out of this two volatility, the 30-day-weighted average is calculated, taken as square root and multiplied by 100 to get the VIX (Options Exchange 2009).

3 Empirical Results on the Information Content of Model-Free Implied Volatility and Hypothesis

Various papers tested the informational efficiency of the mode-free implied volatility, some using the VIX. This section shall give a brief overview of the existing literature and introduce the hypothesis for this paper.

3.1 Review of Empirical Results on Measuring the Information Content of Model-Free Implied Volatility

Firstly, there are numerious papers testing the informational efficiency of the VIX before 2003, for example Blair, S.-H. Poon and Taylor (2001). These papers are however not included in this literature review section, as they do not use the currently applied model-free implied volatility version of the VIX, which was only introduced in 2003.

An early paper that examines the information content of model-free implied volatility is Jiang, George J & Tian (2003). 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 both univariate and encompassing regression analysis. Using option data from the S&P 500 for the model-free implied volatility calculation and 5-min returns to calculate daily realized volatility, they examine monthly non-overlapping samples of a 6-year sample period between June 1988 and December 1994. Their findings are, that the model-free implied volatility subsumes all the information contained in the BS implied volatility and past realized volatility. Another example is the paper of **bakanove2010**. With daily data from oil futures between November 1986 and December 2006, they evaluate the information content of model-free implied volatility in comparison to historic volatility using a monthly frequency. Using regression analysis, they come to the result that implied volatility subsumes the information contained in historical volatility. In contrast to these results is for example Taylor, Yadav and Zhang (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.

Apart from the papers examining informational efficeiency, there are papers using and extending the implied-volatility approach are for ecample **hao2013**, who extend the VIX by deriving the VIX formulas under a risk-neutral valuation relationship, finding that this GARCH implied volatility is significantly lower than the VIX thus interpreting that the GARCH model can not capture the variance premium.

3.2 Research Question and Hypothesis

The aim of this paper is to use the results from Jiang, George J & Tian (2003), 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 index. Thus this paper will examine directly the information content of the VIX index, using a more recent and longer time period of 17 years, between January 2000 and December 2017. Moreover, the approach from the HAR-RV model as described by **Corsi2009** is used.

Model-free implied volatility uses the forward-looking nature of options in comparison to historic volatility and does not suffer from the misspecification problems that occur when testing BS implied volatility,moreover it aggregates information across all strike prices. It is therefore assumed, that it is informationally more efficient than historic volatility. BS implied volatility is not included in the test conducted in this paper, as consistent with previous research it is assumed that for the implied volatility measures, model-free implied volatility outperforms BS implied volatility.

Consistent with previous literature, the following hypothesis are tested:

Hypothesis 1: The VIX contains information about future realized volatility

Hypothesis 2: The VIX is an unbiased estimator of future realized volatility

Hypothesis 3: The VIX has more explanatory power than the historical volatility in estimating future realized volatility

Hypothesis 4: The VIX incorporates all information regarding future realized volatility, historic volatility contains no information beyond what is already included in the VIX

4 Methodology and Data

4.1 Data and Calculation of Input Factors

The data used in this study are from several sources. The realized variance of the S&P 500 index is obtained from the *Oxford-Man Institute's realized library V0.2* (2009). Daily VIX values are the closing values from CBOE. As of September 22, 2003 CBOE started to use a revised methodology for the VIX, but calculated the prices with the new methodology ex-post, dating back until 1990. Therefore the data was taken both from the ex-post Chicago Board of Options Exchange (n.d.[b]) and the daily-updated Chicago Board of Options Exchange (n.d.[c]). The S&P 500 which is only used for visualisation are closing values also taken from Chicago Board of Options Exchange (n.d.[a]). The sample consists of daily data in the period from January 2000 to December 2017.

Firstly, the model needs daily realized volatility, to assess the information content of implied volatilities. The calculation of the daily realized variance is calculated by *Oxford-Man Institute's realized library V0.2* (2009) using the sum of squared 5-min high-frequency returns, as explained in 2.2.1. The formula is given by

$$\sigma_t = \sum x_t^2 \tag{11}$$

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

Previous research has used different sampling frequencies for the calculation of the realized volatility. Whereas earlier studies used mainly daily returns, most recent studies argue in the favor of using intraday returns, as they have certain advantages over daily data (Jiang, George J & Tian 2003). For example Andersen et al. (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 and more importantly, **andersen1998** show that realized volatility calculated using squared returns produces inaccuracies when daily retruns are used. The reaqlized variace data is cleaned by *Oxford-Man Institute's realized library V0.2* (2009) in four ways. First, entries outside the timestamp when exchanges are open are deleted. Secondly, entries with the same time stamp are 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) are deleted. Lastly, entries for which the mid-quote deviated largely from the mean where deleted. To obtain the realized volatility, the realized variance is simply squared

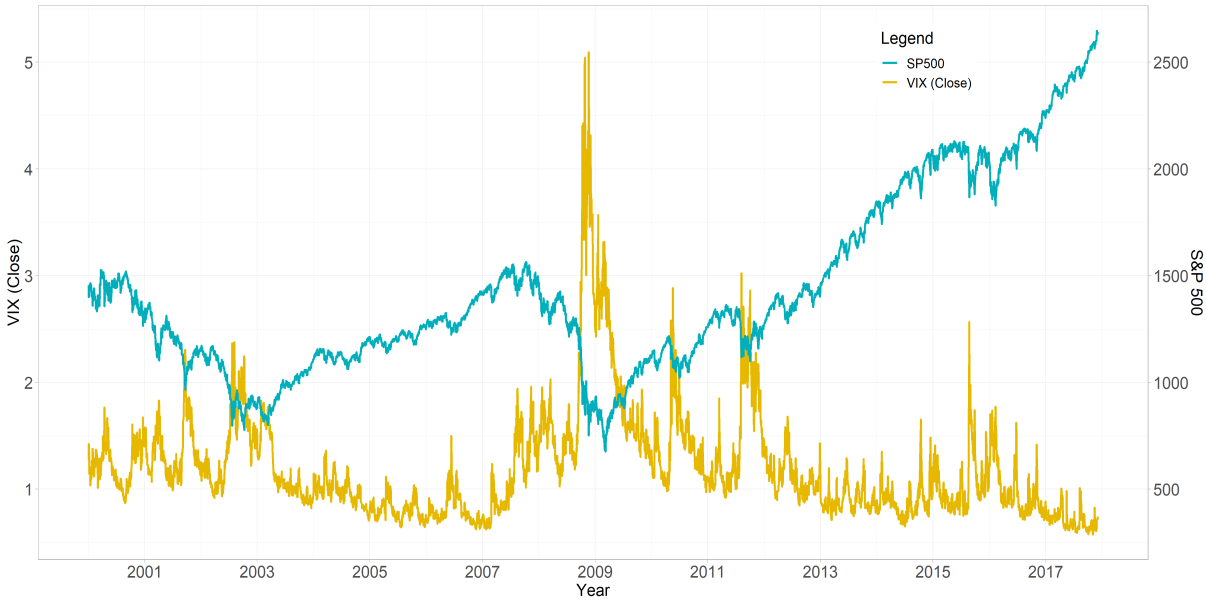
$$RV_t = \sqrt{(\sum x_t^2)}. \tag{12}$$

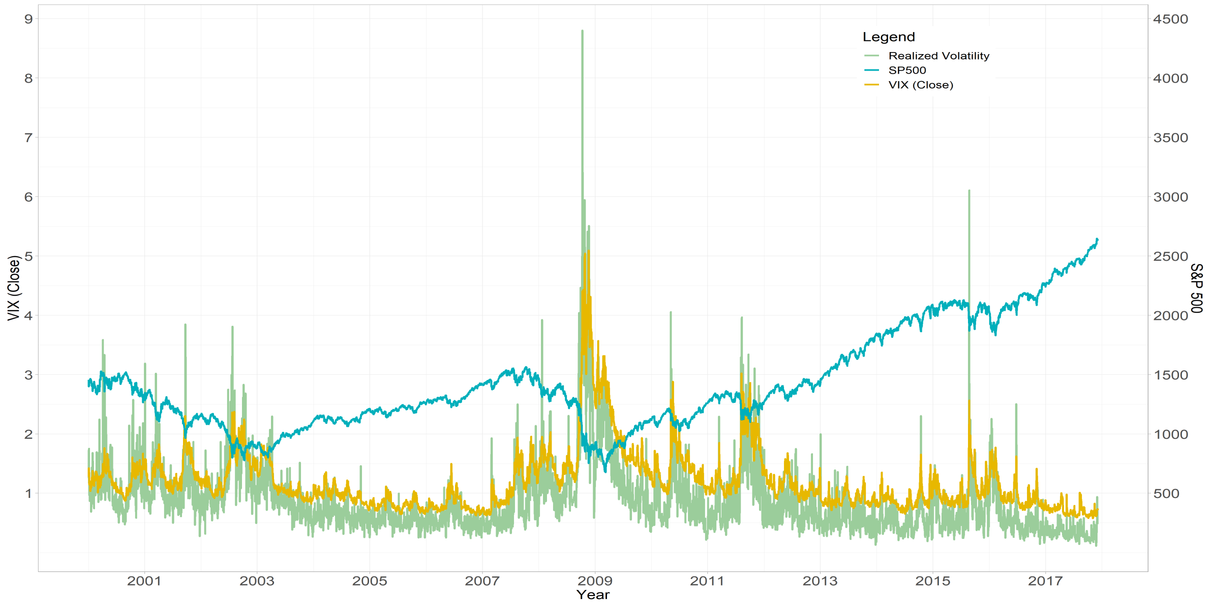
Moreover, the model needs historic volatility, to capture the information contained in the past realized volatility. Therefore simply lagged realized volatiltiy is used. Finally, the VIX is needed to measure the information contained in this model-free implied volatiltiy measure. For the VIX the daily closing value is taken, as it contains the information from the whole day. The calculation is described in 2.3.2. To display the data in a more intuitive manner, the annualized VIX is divided by the square root of 252 as in Blair, S.-H. Poon and Taylor (2001) and Whaley (2008), in order to obtain an index with daily information content.

The S&P 500 together with the VIX is illustrated in 4.1 and the realized vola is added in 4.1. The summary statistics can be found in 9 and **??** in the appendix. The summary statistics show, that the mean of the VIX is for every time period a slightly higher than the mean of the realized variance for all time periods. From the graphics, too, it can be seen that usually the VIX is a little higher than the realized volatility. This is consistent with the findings previous research, e.g. Jiang, George J & Tian (2003). However, during crisis periods, the realized volatility exceeds the VIX. For example, whereas both the realized volatility and the VIX are particularly high during the period of the crisis between 2008 and 2012, the realized variance is significantly higher. To account for this period, a dummy 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 specified and should be statistically better specified than those based on simple volatility.

The correlation matrix for the realized volatility, it's lagged specifications and the model-free implied volatility and it's lags can be found in **??**. Overall, the realized volatility is highly correlated with both the past realized variance and the past VIX values. Whereas for the one day lag the correlation with the VIX is higher, for the weekly and monthly averages the correlation with the realized variance is higher.





4.2 Methodology: Linear Regression and HAR-RV model

Consistent with prior research, for example Jiang, George J & Tian (2003), Canina and Figlewski (1993) or Bent J Christensen and Prabhala (1998), both univariate and encompassing regression analysis is used to analyse the information content of volatility measures. However, contrary to previous research, the approach from the HAR-RV model, described by **Corsi2009** is used. This means, not only one day lagged realized volatility is included as an explanatory variable in the regression, but also weekly and monthly realized volatility. They are computed using simply rolling averages, the weekly volatility is the average over 5 days and the monthly volatility the average over 22 days.

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. 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 volatility one day ahead. The three regressions are then given by:

$$RV_{t+1d} = c + \beta^d RV_t^d + \beta^w RV_t^w + \beta^m RV_t^m \tag{13}$$

$$RV_{t+1d} = c + \beta^{VIX} VIX_t \tag{14}$$

$$RV_{t+1d} = c + \beta^d RV_t^d + \beta^w RV_t^w + \beta^m RV_t^m + \beta^{VIX} VIX \tag{15}$$

with RV_{t+1d} the realized volatiltiy one day ahead, RV^d the daily realized volatility (lagged compared to the explained variable), 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.

In alignment with Corsi (2009), the Newey-West covariance correction is used, to account for the possible presence of serial correlation in the data, as serial correlation causes both the Gauss-Markow and Classical linear model assumptions to fail.

The hypothesis described in 3.2 are then formalized in the following way. For **Hypothesis 1**, if model-free implied volatility contains information about realized volatility, the $H_0 : \beta^{VIX} = 0$ is should be tested in the regression containing only the VIX and the crisis dummy (with the expectation to be rejected). For the **Hypothesis 2**, the $H_0 : \alpha = 0 \; and \; \beta^{VIX} = 1$ should be tested in the same regression (with the expectation to be not rejected). For the **Hypothesis 3**, that the VIX has more explanatory power than the historical volatility in estimating future realized volatility, the adjusted R^2s and AIC should be compared in Regressions containing only the VIX or only historic volatility and the crisis dummy. Finally, for **Hypothesis 4**, that the VIX incorporates all information regarding future realized volatility and that historic volatility contains no incremental information compared to the VIX, the $H_0 : \beta^d = \beta^w = \beta^m = 0 \ and \ \beta^{VIX} = 1$ shall be tested (expected to not be rejected).

4.3 Limitations

As volatility is stochastic, the ex-ante estimation will not equal the return volatility, as it is a measurement over an aggregated discrete time period (**andersen2001**). The VIX might be flawed, as **jiang2007** showed. There might be a bias in estimated realized volatility due to autocorrelation in intraday returns (Jiang, George J & Tian 2003).

5 Results

This section presents the results, obtained with the regression analysis described in section 4.2.

5.1 Regression Analysis Results

Table (1) provides the summary statistics for both the realized volatility and the VIX. It can be seen, that for all time periods the VIX is higher than the

5.2 Robustness Checks

5.2.1 Monthly non-overlapping samples

5.2.2 IV Regression

6 Discussion

A Tables

Table 1: level regression

	Dependent variable:		
	Realized Volatility		
	(1)	(2)	(3)
Intercept	0.045***	−0.324***	−0.167***
	(0.015)	(0.059)	(0.034)
RV^d	0.362***		0.257***
	(0.038)		(0.040)
RV^w	0.391***		0.287***
	(0.056)		(0.064)
RV^m	0.188***		−0.105**
	(0.036)		(0.050)
crisis	0.025*	−0.213***	−0.111***
	(0.013)	(0.035)	(0.021)
VIX		1.051***	0.576***
		(0.059)	(0.064)
AIC	2817.4	3111.9	2450.1
Observations	4,434	4,433	4,433
R^2	0.708	0.688	0.732
Adjusted R^2	0.708	0.688	0.731
Residual Std. Error	0.332 (df = 4429)	0.344 (df = 4430)	0.319 (df = 4427)
Note:			*p<0.1; **p<0.05; ***p<0.01

Table 2: logarithmic regression

	Dependent variable:		
	Realized Volatility		
	(1)	(2)	(3)
Intercept	-0.043^{***}	-0.406^{***}	-0.186^{***}
	(0.007)	(0.017)	(0.014)
RV^{d}_{log}	0.344^{***}		0.264^{***}
	(0.027)		(0.025)
RV^{w}_{log}	0.395^{***}		0.285^{***}
	(0.035)		(0.035)
RV^{m}_{log}	0.208^{***}		0.015
	(0.024)		(0.029)
crisis	0.020^{*}	-0.224^{***}	-0.099^{***}
	(0.012)	(0.034)	(0.016)
VIX_{log}		1.472^{***}	0.644^{***}
		(0.048)	(0.045)
AIC	1874.2	2372.2	1555.5
Observations	4,434	4,433	4,433
R^2	0.726	0.693	0.745
Adjusted R^2	0.726	0.693	0.745
Residual Std. Error	0.299 (df = 4429)	0.316 (df = 4430)	0.288 (df = 4427)
Note:			*p<0.1; **p<0.05; ***p<0.01

Table 3: level regression

	Dependent variable:		
	Realized Volatility		
	(1)	(2)	(3)
Intercept	0.045***	−0.238***	−0.015
	(0.015)	(0.032)	(0.021)
RV^d	0.362***		0.111**
	(0.038)		(0.049)
RV^w	0.391***		0.277***
	(0.056)		(0.063)
RV^m	0.188***		0.364***
	(0.036)		(0.072)
crisis	0.025*	−0.164***	−0.015
	(0.013)	(0.031)	(0.018)
VIX^d		1.562***	1.388***
		(0.110)	(0.125)
VIX^w		0.005	−0.694***
		(0.157)	(0.141)
VIX^m		−0.599***	−0.497***
		(0.131)	(0.114)
AIC	2817.4	2673.3	2033.4
Observations	4,434	4,434	4,434
R^2	0.708	0.718	0.756
Adjusted R^2	0.708	0.717	0.756
Residual Std. Error	0.332 (df = 4429)	0.327 (df = 4429)	0.304 (df = 4426)
Note:			*p<0.1; **p<0.05; ***p<0.01

Table 4: logarithmic regression

	Dependent variable:		
	Realized Volatility		
	(1)	(2)	(3)
Intercept	-0.043^{***}	-0.404^{***}	-0.096^{***}
	(0.007)	(0.016)	(0.011)
RV_{log}^{d}	0.344^{***}		0.126^{***}
	(0.027)		(0.024)
RV_{log}^{w}	0.395^{***}		0.341^{***}
	(0.035)		(0.039)
RV_{log}^{m}	0.208^{***}		0.351^{***}
	(0.024)		(0.040)
crisis	0.020^{*}	-0.195^{***}	-0.021
	(0.012)	(0.033)	(0.013)
$VIX1d_{log}$		2.039^{***}	1.781^{***}
		(0.086)	(0.082)
VIX_{log}^{w}		-0.222^{*}	-1.098^{***}
		(0.121)	(0.118)
VIX_{log}^{m}		-0.412^{***}	-0.468^{***}
		(0.104)	(0.078)
AIC	1874.2	2215.4	1174.6
Observations	4,434	4,434	4,434
R^2	0.726	0.704	0.766
Adjusted R^2	0.726	0.704	0.766
Residual Std. Error	0.299 (df = 4429)	0.310 (df = 4429)	0.276 (df = 4426)
Note:			*p<0.1; **p<0.05; ***p<0.01

Table 5: level regression

	Dependent variable:		
	Realized Volatility		
	(1)	(2)	(3)
Intercept	0.046	−0.088	0.022
	(0.031)	(0.068)	(0.034)
RV^d	0.408***		0.408***
	(0.109)		(0.106)
RV^w	0.501***		0.513***
	(0.126)		(0.127)
RV^m	0.072		0.004
	(0.079)		(0.120)
crisis	−0.022	−0.133*	−0.038
	(0.029)	(0.076)	(0.027)
VIX		0.850***	0.065
		(0.072)	(0.083)
AIC	192.5	531.1	194.4
Observations	456	455	455
R^2	0.757	0.484	0.757
Adjusted R^2	0.754	0.481	0.754
Residual Std. Error	0.297 (df = 451)	0.431 (df = 452)	0.297 (df = 449)
Note:			*p<0.1; **p<0.05; ***p<0.01

Table 6: logarithmic regression

	Dependent variable:		
	Realized Volatility		
	(1)	(2)	(3)
Intercept	-0.001	-0.364^{***}	-0.049^{*}
	(0.018)	(0.026)	(0.028)
RV_{log}^{d}	0.346^{***}		0.352^{***}
	(0.064)		(0.064)
RV_{log}^{w}	0.408^{***}		0.432^{***}
	(0.078)		(0.080)
RV_{log}^{m}	0.171^{***}		0.003
	(0.054)		(0.088)
crisis	-0.017	-0.169^{***}	-0.063^{*}
	(0.027)	(0.062)	(0.034)
VIX_{log}		1.229^{***}	0.236^{**}
		(0.074)	(0.104)
AIC	126.4	415.1	123.9
Observations	456	455	455
R^2	0.748	0.522	0.751
Adjusted R^2	0.746	0.519	0.748
Residual Std. Error	0.276 (df = 451)	0.380 (df = 452)	0.275 (df = 449)
Note:			*p<0.1; **p<0.05; ***p<0.01

Table 7: level regression

	Dependent variable:		
	Realized Volatility		
	(1)	(2)	(3)
Intercept	0.046	-0.270^{***}	-0.037
	(0.031)	(0.051)	(0.039)
RV^d	0.408^{***}		0.274^{**}
	(0.109)		(0.117)
RV^w	0.501^{***}		0.142
	(0.126)		(0.228)
RV^m	0.072		0.357
	(0.079)		(0.222)
crisis	-0.022	-0.210^{***}	-0.065^{***}
	(0.029)	(0.050)	(0.025)
VIX^d		1.328^{***}	0.856^{***}
		(0.238)	(0.170)
VIX^w		0.474^{*}	0.124
		(0.258)	(0.330)
VIX^m		-0.779^{***}	-0.749^{**}
		(0.174)	(0.315)
AIC	192.5	229	137.8
Observations	456	456	456
R^2	0.757	0.736	0.787
Adjusted R^2	0.754	0.734	0.784
Residual Std. Error	0.297 (df = 451)	0.309 (df = 451)	0.278 (df = 448)
Note:			*p<0.1; **p<0.05; ***p<0.01

Table 8: logarithmic regression

	Dependent variable:		
	Realized Volatility		
	(1)	(2)	(3)
Intercept	-0.001	-0.364^{***}	-0.098^{***}
	(0.018)	(0.022)	(0.031)
RV^d_{log}	0.346^{***}		0.172^{***}
	(0.064)		(0.061)
RV^w_{log}	0.408^{***}		0.267^{**}
	(0.078)		(0.103)
RV^m_{log}	0.171^{***}		0.252^{**}
	(0.054)		(0.102)
crisis	-0.017	-0.241^{***}	-0.094^{***}
	(0.027)	(0.045)	(0.035)
$VIX1d_{log}$		1.743^{***}	1.350^{***}
		(0.195)	(0.185)
VIX^w_{log}		0.157	-0.435
		(0.253)	(0.326)
VIX^m_{log}		-0.479^{***}	-0.503^{**}
		(0.155)	(0.206)
AIC	126.4	163.6	66.1
Observations	456	456	456
R^2	0.748	0.727	0.782
Adjusted R^2	0.746	0.724	0.779
Residual Std. Error	0.276 (df = 451)	0.287 (df = 451)	0.257 (df = 448)
Note:			*p<0.1; **p<0.05; ***p<0.01

Table 9: summary statistics

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 10: summary statistics

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 11:

	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

B Some More Appendix Figures

References

Andersen, Torben G et al. (2003). ‘Modeling and forecasting realized volatility’. In: *Econometrica* 71.2, pp. 579–625.

Black, Fischer and Myron Scholes (1973). ‘The pricing of options and corporate liabilities’. In: *Journal of political economy* 81.3, pp. 637–654.

Blair, Bevan J, Ser-Huang Poon and Stephen J Taylor (2001). ‘Modelling S&P 100 volatility: The information content of stock returns’. In: *Journal of banking & finance* 25.9, pp. 1665–1679.

Britten-Jones, Mark and Anthony Neuberger (2000). ‘Option prices, implied price processes, and stochastic volatility’. In: *The Journal of Finance* 55.2, pp. 839–866.

Canina, Linda and Stephen Figlewski (1993). ‘The informational content of implied volatility’. In: *The Review of Financial Studies* 6.3, pp. 659–681.

Chicago Board of Options Exchange (n.d.[a]). *Historic daily prices, year = [2018.* `http://www.cboe.com/products/stock-index-options-spx-rut-msci-ftse/s-p-500-index-options/s-p-500-index/spx-historical-data`.

— (n.d.[b]). *Vix Data for 1990 - 2003, year = [2018.* `http://www.cboe.com/products/vix-index-volatility/vix-options-and-futures/vix-index/vix-historical-data`.

— (n.d.[c]). *VIX Data for 2004 to present, year = [2018.* `http://www.cboe.com/products/vix-index-volatility/vix-options-and-futures/vix-index/vix-historical-data`.

Christensen, Bent J and Nagpurnanand R Prabhala (1998). ‘The relation between implied and realized volatility1’. In: *Journal of financial economics* 50.2, pp. 125–150.

Christensen, Bent Jesper and Charlotte Strunk Hansen (2002). ‘New evidence on the implied-realized volatility relation’. In: *The European Journal of Finance* 8.2, pp. 187–205.

Christensen, Bent Jesper, Charlotte Strunk Hansen and Nagpurnanand R Prabhala (2001). 'The telescoping overlap problem in options data'. In: *Draft Paper*.

Corsi, Fulvio (2009). 'A Simple Approximate Long-Memory Model of Realized Volatility'. In: *Journal of Financial Econometrics* 7.2, pp. 174–196.

Day, Theodore E. and Craig M. Lewis (1992). 'Stock market volatility and the information content of stock index options'. In: *Journal of Econometrics* 52.1-2, pp. 267–287.

Derman, Emanuel and Iraj Kani (1994). 'Riding on a smile'. In: *Risk* 7.2, pp. 32–39.

Dupire, Bruno (1997). 'Pricing and hedging with smiles'. In: *Mathematics of derivative securities* 1.1, pp. 103–111.

Dupire, Bruno et al. (1994). 'Pricing with a smile'. In: *Risk* 7.1, pp. 18–20.

Harvey, Campbell R and Robert E Whaley (1992). 'Market volatility prediction and the efficiency of the S&P 100 index option market'. In: *Journal of Financial Economics* 31.1, pp. 43–73.

Jiang, George J & Tian, Yisong S. (2003). 'Model-Free Implied Volatility and Its Information Content'. In: *The Review of Financial Studies* 18.4.

John, C et al. (2006). *Options, futures, and other derivatives.*

Jorion, Philippe (1995). 'Predicting volatility in the foreign exchange market'. In: *The Journal of Finance* 50.2, pp. 507–528.

Lamoureux, Christopher G and William D Lastrapes (1993). 'Forecasting stock-return variance: Toward an understanding of stochastic implied volatilities'. In: *The Review of Financial Studies* 6.2, pp. 293–326.

Malkiel, Burton G and Eugene F Fama (1970). 'Efficient capital markets: A review of theory and empirical work'. In: *The journal of Finance* 25.2, pp. 383–417.

Müller, Ulrich A et al. (1993). 'Fractals and intrinsic time: A challenge to econometricians'. In: *Unpublished manuscript, Olsen & Associates, Zürich.*

Options Exchange, Chicago Board of (2009). *White Paper Cboe Volatility Index*. Tech. rep. Cboe Exchange, pp. 1–23.

Oxford-Man Institute's realized library V0.2 (2009). Tech. rep. [Online; accessed 31-October-2018].

Poon, Ser-huang and Clive W.J. Granger (2003). 'Forecasting volatility in financial markets : A review'. In: *Journal of Economic Literature* 41.June, pp. 478–539.

Rubinstein, Mark (1994). 'Implied binomial trees'. In: *The Journal of Finance* 49.3, pp. 771–818.

Taylor, Stephen J, Pradeep K Yadav and Yuanyuan Zhang (2010). 'The information content of implied volatilities and model-free volatility expectations: Evidence from options written on individual stocks'. In: *Journal of Banking & Finance* 34.4, pp. 871–881.

Tsay, Ruey S (2005). *Analysis of financial time series*. Vol. 543. John Wiley & Sons.

Whaley, Robert E (2008). 'Understanding VIX'. In: