

The competition between implied volatility and model-based volatility

Evaluating the forecasting performance of GARCH, HAR and implied volatility

Master Thesis

Writer: Freek Engels

Supervisor: Dirk-Jan Janssen

Radboud Universiteit Nijmegen

Abstract

The current study presents a comprehensive research that aims to measure whether model-based forecasted volatility or implied volatility is the best method of determining future volatility. Forecasting horizons of one day, one week and one month will be used. The models that are investigated are the GARCH model, the HAR model and the VIX. The aim of this study is twofold. Firstly, to investigate which of the methods yields the best forecasting results in isolation. Second, to find out whether the different models provide any incremental information on volatility compared to the other models. The results of this study can be divided into two categories. One that compares the absolute performance of the models, and one that compares the relative performance of the models. The results on the absolute comparison between the models show that the HAR model offers the most accurate forecast for the daily and monthly forecast horizon. For the weekly forecasting period, the GARCH model is the best at predicting future volatility. Furthermore, the relative comparison between the models indicate that in almost all the cases, all models have independent information over each other. Only for the daily forecasting horizon, the GARCH model does not have independent information over the other models.

Inhoud

1. Introduction.....	4
2. Overview of the models	6
2.1 Implied volatility: VIX – CBOE Volatility Index.....	6
2.2 GARCH Model – The Generalized Autoregressive Conditional Heteroskedasticity Model.....	7
2.3 HAR-RV Model.....	8
3. Literature review	11
2.2 Hypotheses.....	15
4. Methodology.....	16
4.1.1 GARCH data	16
4.1.2 HAR data.....	18
4.1.3 Implied volatility data.....	19
4.2.1 Obtaining the implied volatility forecasts	19
4.2.2 Obtaining the GARCH forecasts.....	20
4.2.3 Obtaining the HAR forecasts	21
4.3 Research method	22
5. Results	24
5.1 In-sample performance HAR and GARCH.....	24
5.2.1 Daily Mincer-Zarnowitz results.....	25
5.2.2 Weekly Mincer-Zarnowitz results.....	26
5.2.3 Monthly Mincer-Zarnowitz results.....	27
5.3 Forecast encompassing regressions.....	27
6. Discussion and conclusion.....	30
6.1 Hypotheses.....	30
6.2 Comparison existing literature	31
6.3 Strengths and weaknesses	33
6.4 Further research.....	34
Appendix.....	35
References.....	40

1. Introduction

Uncertainty in markets is a topic in finance that is extensively studied in the past. For the prevailing uncertainty in various markets, volatility is a widely used measure. In the stock market, volatility is not directly observed. Hence, in the current literature there have been various methods to measure volatility. A widely known method is taking the natural logarithm of stock returns. It resembles the variation of the returns on certain stocks or an index. More recent literature, however, states that realized variance, which is the intraday squared returns, is a better proxy for real volatility.

Because this uncertainty about the future is not preferred by market participants, numerous studies have tried to estimate the future volatility, considering many approaches. A popular approach is constructing a model for calculating future volatility. One of the best-known models for volatility forecasting is the Generalized Autoregressive conditional Heteroskedastic (GARCH) model introduced by Bollerslev (1986). The GARCH model uses information of historical squared logreturns to forecast volatility. Although the model is widely used, current literature points to several shortcomings of the model. Therefore, Corsi (2009) introduced the Heterogeneous Autoregressive (HAR) model that tries to solve those problems. Corsi based his model on the Heterogeneous Market Hypothesis, which claims that market participants react differently to new information. In contrast with the GARCH model, the HAR model uses realized volatility as proxy for real volatility using the method introduced by Anderson et al (2001). The HAR model should be superior in modeling stylized facts. Other methods try to forecast volatility based on implication derived from the market. Examples of these are volatility indexes on a stock index, such as the CBOE Volatility Index (VIX), which measures the implied volatility of the S&P 500.

It is important to note that there is no unambiguous definition of volatility. A distinction can be made between implied volatility and historical volatility. Where implied volatility is a forward-looking measure calculated from a determined options pricing model, historical volatility is backward-looking and measures the past volatility observed in the market. Volatility indexes present implied volatility while models like GARCH and HAR use historical volatilities to estimate future volatility.

Which method is superior in forecasting volatility is an important topic of debate. The results of various studies show different results on which method is best able to forecast future volatility. For example, Latané and Rendleman (1976) and Beckers (1981) show that implied volatility measures deliver better forecasting results than the forecasted volatility computed by several models based on historical volatility. However, other studies conclude that implied volatility and model-based volatility produce the same results. Canina and Figlewski (1993) and Day and Lewis (1992) are two of those

studies. Since the HAR model was introduced only in 2009, the above studies do not include this model. Later studies do however include the HAR model in their research. Seda (2012) finds that the HAR models is hardly outperformed by any other model.

The current study presents a comprehensive research that aims to measure whether model-based forecasted volatility or implied volatility is the best method of determining future volatility. Forecasting horizons of one day, one week and one month will be used. The two models that are investigated are the GARCH model and HAR model. The GARCH model is chosen because it is widely known and often used as a baseline for comparing forecasted volatilities in the literature. The HAR model is chosen because it is a model that is gaining popularity because of its ability to model real-world behavior of volatility in different markets and makes use of realized volatility to produce forecasts. Data on the S&P 500 will be used to obtain forecast of the S&P 500. Therefore, the measure of implied volatility will be the VIX. In the next chapter, both models and indexes are explained in more detail and relevant literature on the forecasting power of all methods is set out.

The current research contributes to the existing literature by comparing the rather new HAR model to the GARCH model as well as implied volatility. The aim of this study is twofold. Firstly, I want to investigate which of the methods yields the best forecasting results. Hence, all methods will be studied in isolation and their results will be compared. In addition, it is interesting to know whether the other methods still provide additional information on volatility that is not captured by the best forecasting method. Therefore, the second aim of this study is to find out whether the different methods provide any incremental information on volatility compared to the other methods. This leads to the following research question:

How does the predictive power of the HAR model, the GARCH model and the VIX for the volatility of the S&P 500 compare to each other?

The results of this study can be divided into two categories. One that compares the absolute performance of the models, and one that compared the relative performance of the models. The relative performance aims to find out whether models that perform worse than the best model in absolute sense, still contain independent information which seems useful. The results on the absolute comparison between the models show that the HAR model offers the most accurate forecast for the daily and monthly forecast horizon. For the weekly forecasting period, the GARCH model is the best at predicting future volatility. Furthermore, the relative comparison between the models indicate that in almost all the cases, all models have independent information over each other. Only for the daily forecasting horizon, the GARCH model does not have independent information over the other models. The VIX, which does not come out on top in the absolute

performance, in all cases except one does possess independent information that is not present in the forecasts of the other models.

The next chapter consists of an overview and explanation of the models whose forecast ability is measured. The third chapter contains the literature review. Light is shed on the difference between implied volatility and model-based volatility and the ability of implied volatility and the different models to forecast volatility. At the end of the chapter the hypotheses are given. Chapter four consists of an overview of the data and presents the methodology used for answering the research question. Chapter five presents the results and analyzes these results. In chapter six, there is a discussion on the comparison of the results of the current study with the existing literature, and an overview of the strengths and weaknesses of the current study and the conclusion.

2. Overview of the models

This chapter provides an overview of the models investigated by the current research and briefly explains the fundamentals of the models. I start with the VIX, which actually is not a model for predicting volatility but rather a measure for volatility which is implied by the market. Then, the GARCH model and the HAR model and their fundamentals are explained.

2.1 Implied volatility: VIX – CBOE Volatility Index

Where stock indexes measure prices, volatility indexes measure the volatility of an index. The VIX is the first and best-known volatility index and measures the implied volatility of the S&P 500. The VIX was constructed by Whaley (1993) and was introduced in 1993 by the Chicago Board Options Exchange (CBOE). The VIX is created for two reasons. Since market volatility is an important variable for investors, a reliable estimate for the expected short-term stock market volatility needed to be created. Second, the VIX could be used as a basis upon which other derivatives, such as futures and options, can be written. Those derivatives can be used by investors to hedge their portfolio resulting in more flexibility.

Originally, the underlying stock market of the VIX was the OEX index, which consists of 100 stocks of highly important U.S. companies. The level of the VIX was based on the implied volatility of 8 near-the-money options and constructed in such a way that it represents the implied volatility of a hypothetical at-the-money OEX option with 22 trading-days, equivalent to 30 calendar-days, to

expiration. A constant time to expiration is a necessity because implied volatility of options on the OEX tend to be non-constant. Short-term OEX options seem to have a higher implied volatility than long-term OEX options (Fleming et al, 1993). In September 2003, the CBOE changed the calculation of the VIX. Whaley (2008) distinguishes two fundamental reasons for this change. First, the S&P 500 (SPX) became more popular and more frequently used compared to the S&P 100. Therefore, the CBOE decided that the SPX should be the new underlying index upon which the VIX is calculated. The second reason is that besides at-the-money and near-the-money options, out-of-the-money options also exhibit essential information regarding market volatility (Bollen and Whaley, 2004). This is especially true for out-of-the-money put prices. Hence, the CBOE also included out-of-the-money options in the computation of the VIX. Since the introduction of the VIX in 1993, the VIX is extensively studied by numerous researchers. Much of this research focusses on forecasting the VIX. However, less research is conducted on the forecasting power of the VIX.

2.2 GARCH Model – The Generalized Autoregressive Conditional Heteroskedasticity Model

The autoregressive conditional heteroskedasticity (ARCH) model was introduced by Engle (1982). When investigating time series data, the model is used to model a change in variance over time. The model describes the variance of the error term in $t=0$ as function of previous t 's error terms. The ARCH model differentiates between unconditional volatility and conditional volatility. Unconditional volatility is simply the computed average historical volatility. The volatility in every t is treated with equal weight. On the other hand, conditional volatility includes information that is available today. The ARCH model allows the conditional variance to change over time as a function of past errors. The ARCH model is presented in the following equation:

$$\sigma_t^2 = \alpha_0 + \alpha_1 \epsilon_{t-1}^2 + \dots + \alpha_q \epsilon_{t-q}^2 = \alpha_0 + \sum_{i=1}^q \alpha_i \epsilon_{t-i}^2 ,$$

Where

σ^2 = the variance

q = number of lagged squared residuals included

α = the weight attached to each epsilon

Epsilon = the error term, divided in a stochastic piece z and a time dependent standard deviation of σ , so that:

$$\epsilon_t = \sigma_t z_t$$

An extension of the ARCH model is the GARCH model, introduced by Bollerslev (1986). The GARCH model extends the ARCH model by allowing for both a longer memory and a more flexible lag structure. According to Bollerslev, the GARCH model permits a more parsimonious description in many situations. The GARCH model looks as follows:

$$y_t = x_t' b + \epsilon_t$$

$$\epsilon_t | \psi_{t-1} \sim \mathcal{N}(0, \sigma_t^2)$$

$$\sigma_t^2 = \omega + \alpha_1 \epsilon_{t-1}^2 + \dots + \alpha_q \epsilon_{t-q}^2 + \beta_1 \sigma_{t-1}^2 + \dots + \beta_p \sigma_{t-p}^2 = \omega + \sum_{i=1}^q \alpha_i \epsilon_{t-i}^2 + \sum_{i=1}^p \beta_i \sigma_{t-i}^2$$

where

Epsilon = A real-valued discrete-time stochastic process

Alpha = the weight attached to each epsilon

w = long term average

Anchor = Information set of all information through time

p = The number of lagged variances included

q = The number of lagged residual errors included

Where in the ARCH model the conditional volatility is a linear function of past volatilities, the GARCH model allows for lagged conditional volatilities. Hence, the model combines an autoregressive component with a moving average component and computes a weighting average of the long run average volatility (w), the volatility predicted for this period (epsilon) and the new information available for this period (sigma).

2.3 HAR-RV Model

A more recent model for forecasting future volatility is the HAR model introduced by Corsi (2009). Corsi states that the ARCH and GARCH models suffer from two weaknesses. The first weakness is that the models are not able to replicate main empirical features of financial data. The second is that the estimation procedure required are often non trivial. An alternative approach is constructing a proxy for latent volatility. Anderson et al (2001) label this proxy as realized volatility. Daily realized volatility is computed by simply summing intraday squared returns. The theory behind realized volatility is that when intraday returns are frequently sampled, the realized volatility can be made arbitrarily close to

the underlying integrated volatility, which is a natural volatility measure. When volatility is treated as observed instead of latent, its properties can be examined directly. Hence, simpler techniques can be used.

Realized volatility can be captured by the following equation:

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

where

Delta = the intraday return = $1d/M$

$R_{t-j\Delta}$ = the continuously compounded intraday return = $p(t - j\Delta) - p(t - (j + 1)\Delta)$

In forecasting future volatility, the HAR model makes use of high-frequency data. High-frequency data is data that is collected at an extremely fine time scale. An example is transaction by transaction data, which helps understanding market microstructure (Tsay, 2000). High-frequency data may contain detailed information about price series like volume, news, time of the day, time between transactions, etcetera (Andersen, 2000).

The HAR model is based on the idea of the Heterogeneous Market Hypothesis. Müller et al (1993) introduced the Heterogeneous Market Hypothesis. A heterogeneous market is the opposite of a homogeneous market, in which all participants interpret news and react to news in the same manner. Hence, a heterogeneous market recognizes differences in behavior of market participants. There are several characteristics of the Heterogeneous Market Hypothesis:

1. Market participants have different time horizons and dealing frequencies. Differences in dealing frequencies result in different reactions to the same news in the exact same market. Different time horizons mean that actors in the market have different time horizons when trading. The time horizons can be divided in a short-term, medium-term and a long-term component. Each component reacts to news in another way, depending its time horizon and characteristic dealing frequency.
2. In a heterogeneous market, there exists a positive correlation between volatility and market presence. This support empirical findings. Because participants settle for different prices and execute their transactions in different market situations, they create volatility. This is opposite to a homogeneous market, in which there is a negative correlation between

volatility and market presence. The idea in this case is that the more actors there are, the faster the price converges to the real market value.

3. The market is heterogeneous in geographical location.

The HAR model is based mainly on the idea of participants having different time horizons and trading frequencies. For example, daily traders and market makers behave very differently from rather passive traders like central banks and pension funds. All those actors interpret news and react to news differently, depending on their time horizon. Hence, all actors cause different kinds of volatility levels, or components. The original HAR model defines three primary volatility components: short-term horizon corresponding to daily or higher trading frequency, medium-term corresponding to weekly trading frequency and long-term corresponding to a trading frequency of a month or longer. The volatility corresponding to each component is defined as partial volatility.

Given all information above, the HAR model can be constructed and can be described as a cascade of partial volatilities. It has a hierarchical structure in the sense that future volatility of lower components depends on realized volatility of higher components, but not vice versa. Furthermore, there is a simple autoregressive element which means that each partial volatility is dependent on its own value in previous periods. Combining both elements, the HAR model is:

$$\tilde{\sigma}_{t+1m}^{(m)} = c^{(m)} + \phi^{(m)} RV_t^{(m)} + \tilde{\omega}_{t+1m}^{(m)}$$

$$\tilde{\sigma}_{t+1w}^{(w)} = c^{(w)} + \phi^{(w)} RV_t^{(w)} + \gamma^{(w)} E_t[\tilde{\sigma}_{t+1m}^{(m)}] + \tilde{\omega}_{t+1w}^{(w)}$$

$$\tilde{\sigma}_{t+1d}^{(d)} = c^{(d)} + \phi^{(d)} RV_t^{(d)} + \gamma^{(d)} E_t[\tilde{\sigma}_{t+1w}^{(w)}] + \tilde{\omega}_{t+1d}^{(d)}$$

Where $RV_t^{(d)}$, $RV_t^{(w)}$, and $RV_t^{(m)}$ are the daily, weekly and monthly realized volatilities and $\tilde{\omega}_{t+1m}^{(m)}$, $\tilde{\omega}_{t+1w}^{(w)}$ and $\tilde{\omega}_{t+1d}^{(d)}$ are the volatility innovations with a mean of zero.

Note that for the monthly future partial volatility, the highest in the cascade, the structure is simple in the autoregressive form. Subsequently, the weekly future partial volatility depends on its own previous realized and the previous volatility monthly partial realized volatility, and the daily future partial volatility depends on its own previous realized volatility and the previous monthly partial realized volatility and the weekly partial realized volatility. Substituting the monthly and weekly partial volatilities in the daily future partial volatility gives

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

This equation is a three-factor stochastic volatility model in which the factors correspond to the past realized volatilities at different frequencies. Again, this function can be simplified to

$$\sigma_{t+1d}^{(d)} = RV_{t+1d}^{(d)} + \omega_{t+1d}^{(d)}$$

where $\omega_{t+1d}^{(d)}$ is the latent daily volatility measure and the estimation error. In this case, realized volatility is not treated as an error-free measure of latent volatility. However, when this equation is substituted in the previous equation and rearranging, gives

$$RV_{t+1d}^{(d)} = c + \beta^{(d)} RV_t^{(d)} + \beta^{(w)} RV_t^{(w)} + \beta^{(m)} RV_t^{(m)} + \omega_{t+1d}^{(d)}$$

where

$$\omega_{t+1d}^{(d)} = \tilde{\omega}_{t+1d}^{(d)} - \omega_{t+1d}^{(d)}$$

This is the original Heterogeneous Autoregressive model for Realized Volatility (HAR-RV model). The model has an autoregressive structure with the feature that it acknowledges that volatilities can be different depending on the time interval. Hence, the model can describe many stylized facts observed in the economy.

3. Literature review

In finance, a highly debated topic concerns the different methods for calculating future volatility. There are different sorts of volatility. Two basic approaches towards measuring market volatility are computing implied volatility and model-based volatility forecasts, which uses historical data as input. This chapter provides a comparison of different visions presented in the existing literature concerning the difference in forecasting power of implied volatility on the one hand and model-based volatility on the other hand. Furthermore, the difference in forecasting power between the HAR model and the GARCH model are explicated. This chapter is the basis for the hypothesis presented at the end of this chapter.

As the name indicates, implied volatility means the level of volatility which is implied by the current state of the market. This level is based on the price changes in options. Hence, implied volatility is forward looking. Implied volatility is not directly observable. To calculate the implied volatility of an asset, a pricing model for volatility needs to be solved (Canina and Figlewski, 1993). The Black-Scholes model is mostly used to calculate the implied volatility of options. The inputs of the Black-

Scholes model are the stock price, the exercise price, the option price, the time to expiration, the risk-free rate and the implied volatility (Black and Scholes, 1973). By solving for volatility in the model, the implied volatility of an option can be calculated.

Another way of forecasting future volatility is utilizing historical data to predict future volatility. Hence, this is a backward-looking method. In this case, models, such as the GARCH model, are used to determine future volatility by using historical data as input. A problem that arises is that historical volatility often treats volatility as a constant parameter. Therefore, models like the GARCH model, which allow for time-varying volatility, were constructed. However, in those models the parameters of the model themselves need to be constant and accurately estimable (Figlewski, 1994).

Various studies examine the difference in forecasting power between implied volatility and historical volatility. Latané and Rendleman (1976) compute implied volatility by using the basic Black-Scholes model and conclude that implied volatility is highly correlated with actual volatility and that implied volatility is a better predictor of future volatility than standard deviation predictors based on historical data. Beckers (1981) adjusts the concept of Latané and Rendleman for dividend payments and found little evidence that this distorts the predictive power of implied volatility. Other studies conclude that neither of the two methods produce superior results. Day and Lewis (1992) find that implied volatility may contain incremental information relative to conditional volatility from GARCH models. However, they also find evidence that conditional volatility from GARCH models reflect incremental information relative to implied volatility. Lamoureux and Lewis (1993) report similar results to Day and Lewis. Canina and Figlewski (1993) also report similar results in the sense that implied volatility and historical volatility produce the same forecasting results. However, they are more critical because they neither is an appropriate volatility forecaster. Becker et al (2007) examine whether implied volatility measured by the VIX contains information relevant to future volatility beyond the information that is available from model-based volatility forecasts. Their results indicate that implied volatility is not superior in forecasting future volatility. Becker et al (2008) state that, when comparing the various studies concerning the forecasting power of implied and model-based volatility, the general observed pattern is that implied volatility is the best forecast. Some of the authors explain this pattern by relying on the fact that option markets use a wider set of information when forming forecasts. Other authors assign the discrepancy to the difference in the way in which historical data is used across forecasting approaches. Becker et al find that implied volatility does not reflect a wider information set compared to model-based forecasts. This means that both the approaches reflect volatility in the same way.

The above literature mainly compares performance of implied volatility with model-based volatility the ARCH family and older models. More recent literature compares the informational content of implied volatility with more recent models such as the Heterogeneous Autoregressive Model of Realized Volatility. Bush et al (2011) conclude that implied volatility contains additional ex-ante information on volatility beyond that in realized volatility .Fernandes et al (2014) report that is is very difficult for implied volatility indexes to beat the HAR model. This is due to the persisting nature of volatility indexes. A explanation for this is the daily sampling frequency. Because volatility indexes as the VIX reflect the expected volatility of the underlying index of the coming 30 days, looking at daily volatility implied a certain degree of overlapping that worsens the data persistence. Hence, this persistence can become the only feature that matters in forecasting volatility for short forecast horizons. Seda (2012) states that the HAR model is rarely outperformed by any other model, including the GARCH type of models. Vortelinos (2017) compares the four methods for forecasting volatility; Principal Components Combining, Neural Networks, GARCH and HAR. Where HAR comes out on top, GARCH shows the worst results. Moreover, the HAR model produces the most accurate results across all evaluation methods, including R-squared, Root Mean Square Error (RMSE) and Mean Absolute Error (MAE). Huang and Wang (2016) estimate whether the HAR model provides additional information on volatility to the GARCH model by developing a new model in which components of GARCH and HAR are combined. They conclude that introducing the HAR specification better captures the long memory dynamics of volatility. Hence, HAR presents incremental information over GARCH. Będowska-Sójka (2015) study the performance of models using intraday trading and models using daily data for the purpose of Value At Risk (VAR) estimation. Although VAR is not the same as volatility, it also presents a measure of the possible risk attached to certain assets. The conclusion of the paper is that there is no clear answer to the question which model provides the best forecasts.

To the best of my knowledge, there are not many studies that that directly compare forecasting models of the GARCH types, implied volatility and realized variance with each other. A paper that does do this is of Kambouroudis et al (2015). The paper studies the forecasting power of the methods for a number of European and US stock indexes. To access the forecasting power of the competing models, different types of measures are employed. The absolute mean error (MAE) and the root mean squared error (RMSE) are used for an absolute comparison between the different forecasts. Furthermore, a Mincer-Zarnowitz regression, introduced by Mincer and Zarnowitz (1969) is used. In this regression, the dependent variable is the realized variance, which is the intraday squared returns. The independent variable is the forecast of a certain model or volatility index. To assess the relative forecasting performance, a forecast encompassing method introduced by Chong and Hendry

(1986) is conducted. In these regressions, more than one forecasts are included as independent variable to see whether forecasts contain incremental information over each other. This method will also be applied in the current study and more information will be provided in the next chapter. The results of the MAE and RMSE methods are both in favour of the models that make use of realized variance while computing forecasts. Further results indicate that models in which implied volatility is used outperforms all GARCH models. This result is more pronounced in RMSE compared to MAE. Also for the Mincer-Zarnowitz regressions, the realized variance methods outperform the other methods. The R-Squares of the realized variance models almost are almost always the highest. Furthermore, an examination of the relative forecasting performance by the forecast encompassing method shows that all forecast of the different models contain incremental information over each other. This means that the R-Squared of the univariate regression increases when another variable is added and that the estimates for all those forecasts are significant. Hence, even the model with the worst performance in MAE, RMSE and the Mincer-Zarnowitz regression contain independent information that is not present in the other models. Taking into account all different measures, models of realized volatility outperform the GARCH models. Implied volatility always has useful information when used in addition to other models. However, implied volatility in isolation performs worse than realized volatility models and GARCH in general. This results are in line with later research of Kambouroudis et al (2016), which studies the effects of implied volatility and trading volume on forecasting performance of various GARCH models. Their conclusion is that the inclusion of implied volatility and trading volume almost always lead to a better forecasting performance and thus carry independent information.

Another study that directly compares forecasting power of implied volatility, realized variance and the GARCH family is from Kourtis et al (2016). The forecasting power of the models is estimated for a daily, weekly and monthly forecasting horizon for 13 different indexes. They also use the Mincer-Zarnowitz regression and the forecast encompassing regressions as methods to evaluate forecasting power. They find that the forecasting power of the models differs with the forecasting horizon. The univariate regressions show that for the daily forecasting horizon, the HAR model is the most accurate. For the daily forecasting period, the HAR model provides the best forecasts. For the monthly horizon however, implied volatility offers the best forecasts. The results for the weekly horizon are mixed, with different models performing the best for the different indexes. Furthermore, it is notably to state that the results on the S&P 500, also the index used in the current research, show that the GARCH combined with leverage effects offers the best forecasting power for all the three forecast horizons in the univariate regressions. The forecast encompassing regressions also show mixed results. Overall, all models seem to have some incremental information over each other.

However, for some indexes, some models are forecast encompassed by other model. In summary, the daily results suggest that HAR performs the best while the monthly results suggest that implied volatility with a volatility risk premium performs the best. For the weekly horizon it is difficult to present an unambiguously answer on the question which model most accurately forecasts volatility.

Koopman et al (2005) provide a more relative comparison between models of realized volatility, GARCH and implied volatility. Since the HAR model did not exist at the time, they use other models that make use of realized volatility to compare them with GARCH and implied volatility. They report that for the S&P 100, the realized volatility models provide more accurate forecasts compare to GARCH models and implied volatility. These effects are more pronounced when the mean squared error and mean absolute error are used as measure than when the Mincer-Zarnowitz regression is used. Since they do not use the forecast encompassing method or another measure of relative performance, their results present only a absolute performance comparison.

In conclusion it seems that implied volatility is the best method in forecasting future volatility. This is most convincing when comparing implied volatility to the GARCH model. There is less evidence on the comparison of implied volatility and GARCH on the one side and the HAR model on the other side. However, it seems that implied volatility beats HAR and that HAR beats GARCH.

2.2 Hypotheses

In the previous paragraph the forecasting power of model-based volatility and implied volatility was explicated. Most of the literature concludes that implied volatility outperforms GARCH forecasted volatility. There is less evidence on the predictive power of HAR. The existing literature leans to implied volatility as best indicator of future volatility when time horizons are longer but leans to the HAR model when time horizons are lower. Furthermore, HAR seems to outperform GARCH.

Therefore, the hypotheses of this study are as follows:

H1: For a longer time-horizon, implied volatility outperforms both HAR and GARCH in predicting future volatility and HAR outperforms GARCH. For a shorter time-horizon, HAR outperforms both implied volatility and GARCH, and implied volatility outperforms GARCH.

The hypothesis mentioned above aims to compared the absolute forecasting performance of the competing models. However, that the GARCH model would be outperformed by implied volatility and HAR does not mean that the method is useless and does not provide any useful information on future volatility. Many studies state that when one method outperforms another one, the other method can still contain incremental information. The existing literature is inconclusive on this

subject. Therefore, an additional hypothesis about the relative performance of the competing models is included:

H2: Implied volatility, HAR forecasted volatility and GARCH forecasted volatility all contain incremental information on future volatility.

In this study, incremental information is defined as any additional information on future volatility that ameliorates forecasting power of future volatility. In other words, when a combination of two of the methods results in a better fit for forecasting future volatility than a method in isolation, the other method is presumed to possess incremental information. The incremental information can be tested by means of a regression in which all volatilities are included. This is on top of the normal regressions testing for the first hypothesis, which simply regresses a forecasted volatility on the real volatility. In the next chapter the methods are explained in more detail.

4. Methodology

This chapter begins with a description of the data. After this, there will be an explanation of how the forecasts are exactly obtained for each of the models. Finally, an elaboration on the used methodology is given.

4.1.1 GARCH data

For the GARCH model the daily, weekly and monthly forecasts are needed in order to measure the forecasting power of the model. The whole sample period runs from 01-01-2009 until 01-01-2018. The starting date of 01-01-2009 is chosen because the data on the S&P 500 shows that the index shows recovery from the preceding financial crisis. Including the period of the financial crisis in forecasting volatility could disturb the forecasts of the GARCH model. Economically it makes sense to use data from periods of bull and bear markets separately because volatility in bull markets is generally lower than volatility in bear markets. Therefore, data sample is chosen in such a way that it includes just a period in which a bull market is observed. Further research can compare the results obtained in bull markets with results obtained in bear markets. However, this is beyond the scope of the current research.

The data sample is divided into two parts. On the first part of the sample the GARCH model is estimated. The results are then used to forecast volatility for the next period based on the results of this estimation. This part is the out-of-sample part of the data. The period from 01-01-2009 until 01-

01-2016 is used to forecast the period 01-01-2016 until 01-01-2018. The data that is used are the daily closing prices of the S&P 500 which are obtained from the database of Reuters Thompson Eikon. For forecasting, the logreturns are used. Hence, the data is first transformed to logreturns before it is analyzed.

An inspection of the summary statistics of the GARCH data shows a mean and median of respectively 0.0004057 and 0.000586. The standard deviation is 0.0104938. The data has a skewness of -0.3203627 and a Kurtosis of 7.988503. The lowest value is -0.068954 and the highest value is 0.068355. The summary statistics of the GARCH data are presented in table 1.

Before estimating the model and forecasting volatility, the data is tested for heteroskedasticity and autocorrelation. To see whether the data exhibits clustering of volatilities, several tests can be conducted. First, an inspection of the plots of the data is performed. The plots can be found in the appendix. The plot of the daily squared returns of the S&P 500 shows three clusters of peaks in 2009, 2010 and 2011, so it seems the data is characterized by heteroskedasticity. The peaks of 2009 and 2011 are the highest while the peaks in 2010 are moderate. After the 2011 peaks, the data is very stable with the exception of a clustering around the middle of 2015. There is a clustering of reasonable high returns in this period. After this period, the data seems stable again. Figure 1 of the appendix shows the plot of the GARCH data.

To further test for heteroskedasticity, a Breusch-Pagan test can be conducted. The Breusch-Pagan test is described by Breusch and Pagan (1979) and utilizes chi squared to determine whether the data is characterized by heteroskedasticity. Under the null hypothesis, all error variances are equal while under the alternative hypothesis, the error variances are not equal, hence, there is heteroskedasticity. The null hypothesis can be rejected at a significance level of 5%.

To test for autocorrelation, a Ljung Box test, described by Ljung and Box (1978), is used. The Ljung-Box test can be used to test whether the returns on an index are autocorrelated. Under the null hypothesis, there is no autocorrelation while the alternative hypothesis states that there is autocorrelation present. When the alternative hypothesis is accepted, there is an ARCH effect present in the data. The null hypothesis can be rejected at a significance level of 5%.

Another approach for testing the arch effects is the Lagrange multiplier (LM) test on the squared returns described by Engle (1982). The arch LM tests for autocorrelation in the squared returns. Hence, it tests for conditional heteroskedasticity in the data. The LM test has a null hypothesis of no presence of ARCH effects and the alternative hypothesis of presence of arch effects. The null hypotheses can be rejected for 22 lags at a significance level of 1%. The data is characterized by autocorrelation in the squared returns.

Next to an inspection of the return data an inspection of the residuals of the returns is required. Because the GARCH model is used to forecast volatility of the out-of-sample period, the residuals for the in-sample period are required. For this purpose, the return data is first regressed to a constant and the residuals of this regression are obtained. To check whether the residuals are serially correlated, the ARCH LM test on the squared residuals is conducted. Again, the lags of days of one month (22 days) are included in the test. The null hypothesis of no ARCH effects can be rejected at a significance level of 1%.

4.1.2 HAR data

The HAR model utilizes high frequency data to model and forecast volatility. The data is downloaded from realized.oxford-man.ox.ac.uk. The data consists of the 5 minute intraday squared returns of the S&P 500. The plot of those intraday squared returns reveals that the in-sample period is very similar to that of GARCH data, which is to be expected. However, taking a look at the out-of-sample data shows that there is a very large peak in august 2015 in the HAR data, which is not present in the GARCH data. This peak is especially large because of the extremely large observation on the 24th of august, which is almost 10 times higher than that of the previous and next day. When fitting the HAR model on the out of sample data, this observation results in an exceptionally bad fit of the model. The observed R-Squared was 0.13, compared to an R-Squared of 0.532 for the in-sample period. When the observation is excluded from the sample, the R-Squared of the model rises with 0.4, to an R-squared with a reasonable level. Hence, the observation biases the results in such a way that the HAR model performs way worse than would be expected respecting the existing literature. Therefore, in the data of HAR as well as GARCH and VIX, the observations of the 24th of august 2015 are removed from the data. Figure 3 and of the appendix presents the plot of the HAR data with respectively the inclusion and exclusion observation of the 24th of august 2015.

As will be explained in the next section, the daily, weekly and monthly realized volatilities are needed in order to make the forecasts. The daily realized variance has a minimum value of 0.00000121877 and a maximum value of 0.00164505. The mean and median are respectively 0.0000842445 and 0.0000368543. The standard deviation is 0.000140418 and the skewness and kurtosis are respectively 4.64431 and 33.4546. The weekly data has a minimum of 0.0000169179 and a maximum of 0.00601998. The mean and median report values of respectively 0.000420305 and 0.000217419. The standard deviation, skewness and kurtosis are respectively 0.000597454, 3.61026 and 20.3859. The minimum and maximum of the monthly data are 0.00011429 and 0.012855. The mean and median are respectively 0.00182234 and 0.00102298. Furthermore, the standard deviation, skewness and

kurtosis are 0.00224382, 2.74591 and 10.8836 respectively. These summary statistics show that the skewness as well as the Kurtosis decline as the volatility horizon increases. The kurtosis of the daily and monthly data are above three, indicating that the data is leptokurtic.

4.1.3 Implied volatility data

The data for the implied volatility of the S&P 500 again obtained from the Reuters Eikon database for the period of 01-01-2009 to 01-01-2018. The descriptive statistics of the VIX are presented in table 2. The average annualized implied volatility over the next 30 days of the sample period is 18.68 on average. The median is 16.32, which is slightly lower than the average. The smallest observation corresponds with a value of 9.14 while the largest observation is 56.65. An inspection of the time plot of the VIX shows that the largest observation is in 2009 and that the lowest observation is in 2017. A skewness of 1.68 and a kurtosis of 5.93 indicate that the VIX is not normally distributed since the skewness is not close to 0 and the kurtosis is not close to 3. The Jarque-Bera rejects null hypothesis that the VIX is normally distributed with a significance level of 1%.

Figure 3 shows the plot of the VIX. Taking a look at the plot of the VIX the similarity with the S&P 500 stands out. The peaks that can be seen in the VIX index correspond to the peaks seen in the S&P 500. Furthermore, in the periods of low volatility, the VIX is more volatility than the S&P 500. Therefore, it seems that the VIX index produces better forecasting results in high volatility periods than in low volatility periods. However, the conclusion whether the S&P 500 follows the VIX or vice versa cannot be drawn based on the plots.

4.2.1 Obtaining the implied volatility forecasts

The forecasts of the three models are obtained using different methods for every model. The most simple method is that of the VIX. As stated by Whaley (2009), the VIX shows the annualized expected volatility on the S&P 500. However, in order to compare the VIX forecasting performance with GARCH and HAR for the chosen horizons, the daily, weekly and monthly expected volatility are needed. Since volatility is proportional to the square root of time, the level of the volatility index needs to be multiplied by the square root of the ratio of time. Let h be the time period for which the volatility must be calculated. The corresponding level in the S&P 500 for the time period will be:

$$VIX_{h,t} = VIX_t * \sqrt{\frac{h}{252}}$$

When summing the trading days in the sample and dividing them by the years of the sample, the average trading days per year is approximately 252. The corresponding h for one day will be 1, the h for one week will be 5 and the h for one month will be 22. Hence, when the VIX quotes 30 at time t, expected volatilities for respectively one day, one week and one month will be:

$$30 * \sqrt{\frac{1}{252}} \approx 1.89,$$

$$30 * \sqrt{\frac{5}{252}} \approx 4.23,$$

$$30 * \sqrt{\frac{22}{252}} \approx 8.86$$

4.2.2 Obtaining the GARCH forecasts

The data that is used for forecasting volatility with the GARCH model consists of the daily closing prices of the S&P 500 and is obtained from the database of Thompson Reuters Eikon. After obtaining the data, the data is transformed into logreturns. Only the daily logreturns are needed to forecast volatility for every forecast horizon. Next, the GARCH model is fitted on the in-sample period using the program TimesSeriesModeling 4. This statistical program is used because it provides an easy way for forecasting volatility for all models. Because the GARCH (1,1) model is used, one lagged autoregressive order and one lagged moving average order is included in the estimation.

Furthermore, the intercept, presenting the long-run average, is included. When the model is estimated and seems appropriate, the forecasts can be produced. To produce the daily, weekly and monthly forecasts, the program runs daily forecasts for 22 days ahead. This is the only forecast that is made and thus the output is used to construct the forecasts for every horizon. The output consists of the forecasts of the logreturns and the variance of the forecasts of the logreturns for the next 22 days. A rolling method with a rolling window of the size of the in-sample period is then used to forecast the volatility. The forecasts are obtained recursively. The rolling method first uses all observations from the in-sample period to forecast the logreturn and volatility for the next day. Then the first observation is removed from the window and the forecasted value is included in the window to forecast the logreturn and volatility for the second day. This process is repeated until all forecasts for the out-of-sample period are obtained. Since the GARCH model only uses historical data, a larger in-sample period has an advantage because a lot real data is used compared to a small in-sample period. In this study, the last forecast still uses half of the observations from historical data since the in-sample period is twice as large as the out-of-sample period. When the in-sample period would, for

example, be 3 years and the out-of-sample period 6 years, this would mean that the forecast after the third year would be based solely on forecasts. This would be a disadvantage for the GARCH model in comparison with HAR because the latter always uses historical information of realized from the previous day, which intuitively has more right information than that of a forecast. However, a disadvantage of a large in-sample period is that old peaks or trends can still influence the later forecasts because those are still used to forecast volatility. Since a long-run average is included in the GARCH model, this means that this average is updated slowly with a long rolling window. This means that there is a tradeoff between using a large in-sample period versus a small in-sample period when the rolling window has the size of the in-sample period.

After obtaining the variance for the forecasted logreturns for the next 22 days directly from the output, the forecasts can be finalized. For the daily forecasts, the forecasts of the variance of one day ahead are used directly. To obtain the weekly forecasts, the average of the 5 days ahead forecasts are used. Finally, for the monthly forecasts, the average of the 22 days ahead forecasts are used.

4.2.3 Obtaining the HAR forecasts

The forecasts with the HAR model are produced in a different manner. Since HAR uses a simple autoregressive structure, the model can be estimated by OLS. In contrast with GARCH, which estimates one model and makes one forecast to produce the forecast for every horizon, the HAR model estimates three models which are used separately to estimate the forecasts for every period. The difference between the three models for every forecast horizon is just the dependent variable. For the daily forecast period, the dependent variable is the daily realized variance. For the weekly and monthly forecast period, the dependent variables are respectively the weekly and monthly variances. To obtain the weekly realized variances, the daily realized variances of the past 5 days are summed together and divided by 5. Similarly, to obtain the monthly realized variances, the daily realized variances of the past 22 days are summed together and divided by 22. The independent variables in all three models are the lagged values of the daily, weekly and monthly realized variances. For the estimation of each of the models, those independent variables do not change.

After estimation of all three models with the OLS regressions, the forecasts can be made. Where the GARCH model produces both forecasts and corresponding coefficients, the HAR model only forecasts the coefficients. For the GARCH model the forecasted volatilities are directly produced. To compute the final volatility forecasts with the HAR model, the forecasted coefficients must be multiplied with the lagged realized variances. Since there are three lagged variables included, the model forecasts three coefficients. After multiplying the coefficients with the lagged variables, the outcomes are

summed together to arrive at the forecasts. The final forecast also includes a intercept. However, for all forecast horizons these intercepts turn out to be zero.

4.3 Research method

The research question will be answered by performing a horserace on the S&P 500 between the forecasted volatilities calculated by the HAR model, the GARCH and implied volatility which is given by the VIX. The methods for measuring the forecast performance of the models are divided into two categories. The first category aims to measure absolute forecasting performance. In this case, there will be a ranking in the performance of the models. One model will be the best, one model second best and one model will be the worst forecaster. The second category aims at measuring the relative forecast performance of the models. When a model shows the best absolute forecast ability, this namely does not mean that the forecasts of the other models are worthless. Moreover, it could be the case that there is more independent information present in the model with comes in third than in the model that comes in second when those are compared to the model with the best forecasts.

After computing all forecasted volatilities, first the absolute performance of the models is tested. A regression analysis is performed in which the true volatility is the dependent variable and in which the forecasted volatilities and the implied volatilities are the independent variables, following the testing procedure of Mincer and Zarnowitz (1969). Hence, there are three regressions in the following form:

$$\sigma_t^2 = \alpha + \beta h_t^2 + \varepsilon_t$$

where

σ_t^2 = a measure of true volatility

h = the forecasted volatility of HAR, GARCH or VIX

The Mincer and Zarnowitz (MV) regression measures how much of the true volatility can be explained by each of the forecasted volatilities. The R-squares of the regressions are compared and the beta with the highest R-squared is the best forecast. The coefficient of the forecast must be significant, where a significance level of 95% is used.

Besides the Mincer-Zarnowitz regressions the root mean square error (RMSE) is used as measure to asses the relative performance of the competing models. The root mean square error is the measures the standard deviation of the residuals of a model and is frequently used as a measure of forecasting performance. It measures the differences between the predicted values of a model with

the values which are actually observed. Since RMSE measures standard deviation, the value should always be above zero. A value of zero represents a perfect fit of the forecast with the real data. The higher the value becomes, the worse the fit is.

To measure the relative forecasting performance of the models, the forecast encompassing method introduced by Chong and Hendry (1986) and later modified by Fair and Shiller (1988) is used to examine the incremental value of each method in comparison to the other methods. Here, four regressions are conducted to see whether all models contain independent information which is not present in one or both other models. In the first regression, all three forecasted volatilities are included as independent variables and the dependent variable, again, is the true volatility. The corresponding regression model is

$$\sigma_t^2 = a_0 + a_1 h_{1,t}^{2f} + a_2 h_{2,t}^{2f} + a_3 h_{3,t}^{2f} + \epsilon_t$$

where h_1 , h_2 and h_3 are respectively the forecasted volatility of the HAR model, the forecasted volatility of the GARCH model and the VIX. In the remaining three regressions, only two of the forecasts are included as independent variables.

The forecast encompassing method investigates the relative performance of HAR, GARCH and implied volatility. When one of the models contains no incremental information on future volatility, one or both the other models are encompassing the concerning model. For example, if GARCH shows no additional information relative to HAR and implied volatility, one or both of the latter two encompasses GARCH. In this instance, the corresponding coefficient is statistically not different from zero. This confirms the null hypothesis that the coefficient is zero. To conclude that a model contains incremental information over another model, the R-Squared of the model should increase when the forecast of another model is added to the univariate regression as second independent variable. At the same time, the estimations of the model should be significant different from zero. When this is not the case but the R-Squared still increases, one cannot state that the model has incremental information. Also for the forecast encompassing method, a significance level of 95% is used.

The estimation of the absolute as well as the relative forecast performance of the models is done using different time horizons. The time horizons that are used are one day, one week and one month. The daily forecast horizon will be exploit to answer the question which model provides the most accurate forecast for short-term volatility, where the monthly forecast horizon will be exploit to answer the question which model provides the best long-term forecasts. The weekly horizon lies in between the two and can be seen as a middle-term horizon. Furthermore, only trading days are included. Hence, the weekly horizon consist of 5 days and the monthly horizons consists of 22 days.

5. Results

In this chapter the results of the research are presented. In the next section, the in-sample performance of the HAR and GARCH model are presented. Next, there is an overview of the daily, weekly and monthly absolute performance of the three models. In the last section, the relative performance is assessed using the results of the forecast encompassing regressions.

5.1 In-sample performance HAR and GARCH

In this section the in-sample performance of the HAR and GARCH models is presented. For the VIX it is not necessary to analyze the in-sample period since the forecasts are directly observable and therefore directly compared with the realized variance. However, with the HAR model the estimates are forecasted and with the GARCH model even the volatilities itself are forecasted. Hence, it makes sense to take a close look at the data of the in-sample period to assess the quality of the data.

Since the GARCH model only uses the historical daily squared returns to forecast volatility, only one model has to be analyzed. This is done by fitting the model on the data from the in-sample period. For this purpose the Gaussian regression with a maximum likelihood estimation is used. After estimating the model, the forecasts are obtained using a fixed rolling window as explained in the previous chapter. The GARCH model has three components, which are the intercept, the alpha and the beta. Together, the alpha, beta and the coefficient corresponding to the long term average should be lower or equal to zero. The estimation of the model resulted in an intercept of 0.00161, an alpha of 0.1137 and a beta of 0.86322. The coefficients added together are below 1. All values are statistically significant. Furthermore, a Jarque-Bera test is performed, which give a statistic of 150.178. The residual skewness and Kurtosis are respectively 0.5648 and 4.0555. Finally, the Box-Pierce test on the residuals is 0.6356. A summary of the above can be found in table 3 of the appendix.

To measure the fit of the HAR model on the in-sample data, a simple OLS regression is performed since there is an autoregressive structure. Three models have to be estimated in which only the dependent variable changes in daily realized variance, weekly realized variance and monthly realized variance. The independent variables always are the lagged values of the daily, weekly and monthly realized variances. Hence, for every model there are three coefficients. In the original paper of Corsi (2009) there is also an intercept included. However, the intercepts of all three models are statistically zero. For the daily model, the coefficients are 0.35543, 0.23835 and 0.34522 for respectively the

daily, weekly and monthly lagged realized variances. All coefficient are statistically significant. What we see here is that the weekly realized variance is the weakest explainer of the daily variance of the next day. The R-Squared of the model is 0.532. Hence, the model seems to have a fairly good fit. The weekly model shows coefficients of 0.81199, 0.80085 and 0.00565 for respectively the daily, weekly and monthly lagged variances. In this case, only the daily and weekly lagged variances are statistically significant. The monthly lagged variance does not seem to have a meaningful effect on the weekly variance. The R-Squared of 0.9566 is much higher than that of the daily forecasting period. This means that the HAR model seems a better fit when forecasting weekly volatility. Finally, for the monthly model the coefficients of the daily, weekly and monthly lagged variances are 0.26683, 0.22296 and 0.93663. All coefficients are statistically significant. In contrast with the weekly model, here the monthly lagged realized variance has a large effect on the monthly variance. The R-Squared of the monthly model is 0.9874, which is the highest of all three models. Since also all coefficients are significant, the monthly model seems promising. Table 4 gives an overview of the coefficients and R-Squares for all forecast horizons.

5.2.1 Daily Mincer-Zarnowitz results

In this section the performance of the models based on the daily forecasts is evaluated. The Mincer-Zarnowitz regressions on the forecasts of the S&P 500 include the intraday squared returns as dependent variable and the forecasts of the models as independent variable. The HAR model performs the best with a R-Squared of 0.5176, which is just below the R-Squared of 0.532 from the in-sample period. Closely behind the HAR model is the VIX with an R-Squared of 0.5128. The GARCH model clearly performs the worst with an R-Squared of 0.4159. The results of the daily forecast horizon are presented in the table below. All coefficients are significant, hence, this is not reported in the table. An inspection of the RMSE shows that all models perform equally well for the daily horizon. Because of this it is not possible to discriminate between the forecasting power of the models using this measure.

	GARCH	VIX	HAR
R-Squared	0.4159	0.5128	0.5176
Coefficient (Significance)	0.74624 (0.000)	0.0002 (0.000)	0.65733 (0.000)
Standard deviation	0.11223	2e-005	0.08865

5.2.2 Weekly Mincer-Zarnowitz results

The weekly univariate regression show very different results compared to the daily regressions. For the weekly regressions, the GARCH model has the highest R-Squared of 0.5817. The HAR model comes second with 0.5172. The R-Squared of the regression on the VIX show an R-Squared of 0.5002. What is striking is the fact that the R-Squares of both HAR and VIX decreased slightly. However, the GARCH model shows an increase of 0.1658, which tells us that GARCH is a clearly better forecaster of weekly volatility compared to daily volatility. As with the daily results, all coefficients are significant. The table below presents the results of the weekly forecast horizon.

	GARCH	VIX	HAR
R-Squared	0.5817	0.5002	0.5172
Coefficient (Significance)	6.13501 (0.000)	0.00056 (0.000)	1.13864 (0.000)
Standard deviation	0.92307	8e-005	0.15166

It goes against expectations that the GARCH model outperforms HAR as well as VIX because most literature reports that the HAR model outperforms GARCH. A possible explanation for the higher R-Squared of GARCH in comparison to HAR can be derived from the analyzes of the in-sample data from the previous section. Although the R-Squared from the in-sample period is very high compared to that of the daily forecasting period, the coefficients of the monthly lagged realized variance is insignificant. This could partially nullify the higher R-Squared, leading to an almost identical R-Squared for the daily and weekly forecasting period. This could be the explanation of the fact that GARCH outperforms HAR. However, taking a look at the R-Squares of the in-sample period shows that the R-Squared of the weekly model is much higher than that of the daily model. The question arises which of the two has the largest effect on the forecast performance. Furthermore, the RMSE shows the highest value for the VIX of 0.0004, indicating that the VIX has the worst fit when RMSE is used as a measure of absolute forecasting performance. GARCH and HAR report the same RMSE of 0.0003.

5.2.3 Monthly Mincer-Zarnowitz results

For the monthly forecasting period, the results again show a different order in the models' forecasting performance. As in the daily forecasting period, the HAR model performs the best with a R-Squared of 0.8052. This is the highest R-Squared of all the regressions. Unlike the results of the daily regressions, the gap between the HAR model and the second best performing model is almost 0.1. GARCH shows an R-Squared of 0.7055. Hence, in contrast with the daily regressions, GARCH performs better than VIX, which has an R-Squared Of 0.6079. Again, all coefficients are significant. The results are presented the table below.

	GARCH	VIX	HAR
R-Squared	0.7055	0.6079	0.8052
Coefficient (Significance)	27.7095 (0.000)	0.00089 (0.000)	1.22825 (0.000)
Standard deviation	1.66569	5e-005	0.05637

There are several implications to derive from these results. First, all competing models have by far the highest R-Squared in the regressions of the monthly forecasting period. While VIX and GARCH show an improvement in R-Squared of respectively 0.0951 and 0.1238 compared to their second-best performance, the HAR model even shows an increase of 0.2876. An explanation for the improvement of the VIX could be that the VIX is a forecast for the monthly volatility of the S&P 500 (Whaley, 2009). A second implication of the results is that the gaps in R-Squared between the models is the highest for the monthly forecasting period. The gap between HAR and VIX is 0.1973. The gaps between the best and worst performing models for the daily and weekly forecasting period are respectively 0.1017 and 0.0815. Hence, although the forecasting performance based on monthly data is the best across all models, the variability across forecasting performance also increases. An inspection of the RMSE confirms the results of the Mincer-Zarnowitz regressions. The HAR model shows the lowest value of 0.0006, followed by GARCH with a value of 0.0008. The VIX reports the highest value of 0.0009.

5.3 Forecast encompassing regressions

The purpose of the forecast encompassing regression is to determine whether the methods with a lower R-squared in the Mincer-Zarkowitz regression still contains incremental information over the

method with the highest R-squared. In other words, when the introduction of another variable with a lower R-squared leads to a higher R-squared, the variable with the lower R-squared apparently has some information that is not present in the variable with the higher R-squared. However, for this statement to be true, the corresponding parameters should be statistically significant from zero. To explain the above, we take a look at the daily forecast encompassing regressions. The daily univariate Mincer-Zarnowitz regressions show that the HAR model produces the best forecast with a R-Squared of 0.5176. The goal of the forecast encompassing regressions is to find out whether GARCH and VIX still possess information that is not present in the HAR forecasts. Therefore, all three forecasts are now included as independent variable. The dependent variable is still the daily realized variance. With the inclusion of all three forecasts, the R-Squared of the model rises to 0.5873. Hence, the forecast performance seems to have improved compared to the univariate regression with only HAR included. However, we first have to determine whether the results are statistically significant. For HAR and VIX, this is the case. This means that HAR has incremental information over VIX and that VIX has incremental information over HAR. But when the GARCH forecast is evaluated, it turns out that this result is not significant. In this case it is said that HAR or VIX or HAR and VIX combined forecast encompass the GARCH model. To see which of those options is the case, two other regressions are obtained. In the regression in which only HAR and GARCH are included, the R-Squared becomes 0.5267. This is higher than the univariate regression with only HAR, however, GARCH is not significant. We now know for sure that GARCH does not have incremental information over HAR and can conclude that HAR forecast encompasses GARCH. To see whether also VIX forecast encompasses GARCH, both forecasts are included as independent variable in the regression. Although the R-Squared is again higher than both univariate regressions, GARCH is again not significant. Thus, also VIX forecast encompasses GARCH for the daily forecast horizon. An overview of the results can be found in table 5.

The weekly univariate regressions showed that the GARCH model outperforms both HAR and VIX. The corresponding R-Squared of the regression is 0.5817. The R-Squared of the forecast encompassing regression with all three variables included increases to 0.6171. However, both HAR and VIX turn out to be insignificant. This could mean that GARCH forecast encompasses both HAR and VIX, but this need not be the case. Namely, when we take a look at the regression in which only GARCH and VIX are included, the R-Squared becomes 0.6136 and both GARCH and VIX are significant. This means that VIX in fact does possess information over GARCH, and vice versa. This seems counterintuitive because VIX was insignificant in the regression with all three variables included. The explanation for this lies in the remaining two regressions. From the regression in which HAR and VIX are included it becomes clear that HAR forecast encompasses VIX. Hence, although VIX does have

incremental information over GARCH, it does not have incremental information over HAR. The remaining regression with GARCH and HAR as independent variable shows that GARCH is significant but that HAR is not. Thus, GARCH forecast encompasses HAR. Taking into account all 4 regressions, it is possible to now arrive at a complete story. It has become clear that HAR does not have incremental information over GARCH, that VIX does not have incremental over HAR, but that VIX does have incremental information over GARCH. The fact that VIX does not have incremental information over HAR results VIX to be insignificant in the regression with all three variables. However, a closer inspection has showed that this is only the case because VIX does not have incremental information over HAR. Table 6 presents an overview of the results.

For the monthly forecasting horizon, the univariate regressions show that the HAR model offers the most accurate forecasts with an R-Squared of 0.8052. First, the forecast regression with all three variables is performed. With the inclusion of all variables the R-Squared rises to 0.8746. HAR and GARCH are both significant, VIX is not. When we take a look at the regression with GARCH and VIX, we can see that both GARCH and VIX are significant, so VIX does have incremental information over GARCH. The corresponding R-Squared is 0.7421, which is higher than the R-Squares of the univariate regressions of GARCH and VIX. In the regression with HAR and VIX, also both variables are significant. The R-Squared is 0.8238, again higher than in the univariate regressions. Hence, VIX also has incremental information over HAR. The interesting thing about these results is that VIX apparently does possess incremental information when compared to HAR and GARCH in isolation. But when HAR and GARCH are combined, VIX does not have incremental information over the two anymore. This means that HAR and GARCH must also have incremental information over each other that apparently is the same or more as the incremental information that VIX has. To check this statement, we have to take a look at the last regression. The R-Squared of this regression is 0.8746 and both HAR and GARCH are significant. This means that indeed, both forecasts contain incremental information over each other. The results of the monthly forecast encompassing are presented in table 7 of the appendix.

In conclusion, in most cases all models seem to have incremental information over at least one of the other models. The only exception is the GARCH model in the daily forecast horizon. There, in every of the three regressions in which GARCH is included, it is not significant. For the monthly forecasting period, all models have incremental information over each other. Only VIX does not have incremental information over HAR and GARCH together. The weekly forecast horizon presents mixed results.

6. Discussion and conclusion

This chapter contains the discussion on the results and offers a conclusion. First, the conclusion on whether the hypotheses are accepted or rejected are given. In the following section, the results of this study are compared to the results in the existing literature. After that the strength and weaknesses of the current study are discussed. At the end of the chapter there is a conclusion and recommendations for further research are given.

6.1 Hypotheses

The two hypotheses of this research are as follows:

H1: For a longer time-horizon, implied volatility outperforms both HAR and GARCH in predicting future volatility and HAR outperforms GARCH. For a shorter time-horizon, HAR outperforms both implied volatility and GARCH, and implied volatility outperforms GARCH.

H2: Implied volatility, HAR forecasted volatility and GARCH forecasted volatility all contain incremental information on future volatility.

The first hypothesis cannot be accepted entirely. The expectation that implied volatility outperforms both HAR and GARCH proves not to be true. HAR as well as GARCH clearly outperform VIX for the monthly forecast horizon. However, the expectation that HAR outperforms GARCH turns out to be right. The second part of the hypothesis has proved to be true. HAR offer the most accurate forecasts, closely followed by VIX. GARCH clearly performs the worst.

The second hypothesis is about the relative forecast performance of the models. The hypothesis can partially be accepted. Only for the daily forecast horizon, the GARCH model does not have any incremental information over HAR and VIX. Although HAR provides the best forecasts, VIX does still contain incremental information over HAR. For the weekly forecasting period all models have incremental information over at least one of the other models. GARCH shows the best performance and has incremental information over both VIX and HAR. VIX has incremental information over GARCH, where HAR does not, but in turn HAR has incremental information over VIX. For the monthly forecasting period, all models have incremental information over each other. However, VIX does not have incremental information over HAR and GARCH combined.

6.2 Comparison existing literature

The current study shows similarities as well as differences with the existing literature on forecasting volatility. The most striking result of the current study is that for the weekly forecasting horizon, the GARCH model provides the most accurate forecasts. Most studies that compare the forecasting power of models that use realized variance versus the GARCH family conclude that realized variance models have the greatest forecasting power. For example, Kambouroudis et al (2015) conclude that realized variance models are superior in forecasting volatility in all situations. However, this study does not specifically include the HAR model in their study. The study of Kourtis et al (2016) also compared models of realized variance to various GARCH models. In contrast with Kambouroudis et al (2015), this study does include specifically the HAR model. For most investigated indexes, the HAR model produces the most accurate results for the daily forecasting horizon. However, for the weekly forecasting horizon, the results are mixed, with HAR as well as GARCH models performing the best. Their results on the S&P 500 show that for the weekly forecasting period, the GARCH model has the greatest forecasting power. This seems in line with the findings of this research. However, the GARCH model that is investigated is the GJR-GARCH models, which is a GARCH type model that includes leverage effects. Moreover, they also find that for the S&P 500, the GJR-GARCH model performs the best in the daily and monthly forecasting horizon. This is not in line with the findings of the current research since for the daily period, GARCH performs the worst and for the monthly period, GARCH comes in second. Furthermore, Koopman et al (2005) report that model of realized volatility outperforms GARCH models. However, they only forecast daily volatility and not longer forecast horizons. They also do not use specifically the HAR model, but other models that make use of realized volatility.

Another striking result of the results is that implied volatility performs worse in both the weekly and monthly forecasting horizon. Various studies indicate that implied volatility is a better forecaster than models that use historical volatility. Latané and Rendleman (1976) and Beckers (1981) are two examples of those studies. A footnote on those studies is that the two are relatively old studies produced in times in which high frequency data was not yet available. Hence, intraday squared returns could not be computed at the time. Nowadays, intraday squared returns are mostly used as proxy for real volatility and used as dependent variables in the Mincer-Zarnowitz and the forecast encompassing regressions. In the older studies, often the squared returns are used. Newer studies that do use intraday squared returns report different results. Kambouroudis et al (2015) report that VIX in isolation does not beat HAR and GARCH in forecasting volatility. However, they find that VIX does always contain incremental information over both HAR and GARCH. Kambouroudis et al (2016)

report similar results. The results of the current study also indicate this, with the exception that VIX does not contain incremental information over HAR in the weekly forecasting horizon. In all other cases, VIX does have independent information. Kourtis et al (2016) report very different results in comparison with the current study. Where implied volatility performs the worst in the monthly forecasting period in this study, they conclude that implied volatility offers the best forecasts in this horizon. However, a difference between the studies is that this study incorporates VIX and that Kourtis et al use model free implied volatility which is adjusted for the volatility risk premium. Fernandes et al (2014) report that it is difficult for VIX to beat HAR in forecasting volatility because of the very persistent nature of the VIX. Furthermore, Day and Lewis (1992), Lamoureux and Lewis (1993) and Canina and Figlewski (1993) find that neither model-based volatility or implied volatility produces superior results. They all find that both contain incremental information over each other. The drawback of these studies is again that they are old and thus do not include models which use intraday squared returns as the HAR model.

The results of the current study for the forecast encompassing method indicate that all models have incremental information over at least one of the other models for every forecast horizon. This is in line with Kambourouris et al (2015). They report that linear combination of realized volatility, GARCH and implied volatility are always more informative than the univariate regressions. This is also the case in the current study. A difference is that they also report that a linear combination of the three models is better than all bivariate mixtures of forecasts. However, the results of this research show that for the monthly forecast horizon, the R-Squared of regression with all forecasts included is exactly the same as the bivariate regression with just HAR and GARCH included. Kourtis et al (2016) also report that for the daily and weekly forecast horizon all forecasts contain independent information. In comparison with the current study, they report higher R-Squares. The range of R-Squares they report are 0.707 to 0.727 for the daily horizon and 0.742 to 0.750 for the weekly horizon. The current study reports a range of 0.5267 to 0.5873 for the daily horizon and 0.574 to 0.6171 for the weekly horizon. As stated above, their results on the monthly horizon show that implied volatility is superior in forecasting volatility. Results are mixed in this case, with not all models containing incremental information over other models for various indexes. Furthermore, their R-Squares for the monthly horizon are in turn lower than the R-Squares found in the current study. Another difference between the studies is that Kourtis et al just uses bivariate regressions and do not include regressions with the forecasts of all the models. This limits their ability to investigate whether a model that contains incremental information over two other models in the bivariate regression also contains incremental information over the two other models combined.

In conclusion, I can say that there are similarities as well as differences between the current study and the existing literature. A close inspection of the existing literature tells us that besides differences in results, there are many differences in methodology. For example, older studies mostly used squared returns as proxy for real volatility where newer studies use realized variance as proxy for real volatility. Furthermore, studies differ in forecast horizon, use different measures for forecasting performance and use different lengths of in-sample and out-of-sample periods. Therefore, the results of all studies should be compared carefully and with cautiousness. Overall, the striking results of the current research do to a certain degree confirm the results of other studies. However, in turns there are also studies that do not confirm the results. Hence, we can conclude that there prevails still a lot of disagreement in the topic of the forecasting power of volatility.

6.3 Strengths and weaknesses

The main strength of this study compared to the existing literature is the usage of the forecast encompassing method. Fair and Shiller (1988) explain in what sense the forecast encompassing method introduced by Chong and Hendry (1986) is superior to other methods that try to measure forecasting performance. They state that the forecast encompassing method is a useful alternative to more standard procedures as comparing the size of the root mean squared errors (RMSE's) and mean absolute errors (MAE's), arguing that the first approach is often able to discriminate better. For example, when the difference of RMSE's and MAE's of different forecast models are very close to each other, it is difficult to assess whether these differences are economically meaningful. In other words, because there is only a comparison on the absolute level, it is difficult, if not impossible, to compare different forecasts. Introducing the forecast encompassing method allows for a relative comparison between two forecasts. Although just combining two forecasts with similar RMSE's and MAE's in one regression often does not result in one forecast encompassing the other, combining both of the two separately but with a third forecast can result in one of the forecasts encompassing the third, but the other not forecasting the third. In this sense, the forecast encompassing method is superior to just comparing RMSE's, MAE's or R-squares of the simple Mincer-Zarnowitz regression of different forecasts. Clements and Hendry (1993) state similar shortcomings of comparing MSFE's of different forecasts and also proposes forecast encompassing methods to correct for those shortcomings.

A weakness of this study is that only the S&P 500 is forecasted by various models. Other studies aim to forecast more indexes, and find different results between those indexes. For example, Kourtis et al

(2016) included 13 indexes in their research. For some indexes, they find very different results, which can lead to a more in-depth analyses of the forecast performance of the models.

Another weakness of this research is that only the most basic models are used to forecast volatility. Other studies have showed that the basic HAR and GARCH models which include other effect perform better. For example, Huang et al (2013) report that the HAR model with continuous volatility and jumps and influence of momentum effects performs better than the original HAR model. Corsi and Reno (2009) also state the adding continuous volatility, leverage and jumps play a significant role in forecasting volatility and that neglecting one of them is detrimental to the forecasting performance. For the GARCH model, Glosten et al (1993) introduces the GJR-GARCH model. Brownlees et al (2012) report that the GJR-GARCH model is superior in forecasting volatility compared to other GARCH models. The use of those superior extended HAR and GARCH models could have altered the results of the current study.

6.4 Further research

Taking into account the weaknesses of the current study, future research could focus on expanding the evaluation of forecasting performance across models to other indexes. Especially indexes of less rich countries in for example Africa or South-America are not studied heavily yet. The inclusion of the indexes of these countries enriches literature on the topic. Furthermore, less wide-knowns indexes of the richer countries should be included. The current literature mainly focusses on the countries index with the companies with the highest value. It is interesting to understand whether the results can be translated to indexes with companies that include less known companies. Further research should also focus on comparing the extended models of HAR and GARCH. Although this is already done by some studies, many of these studies just compare the absolute performance of these models. Including a measure of relative performance, such as the forecast encompassing method, could lead to a more comprehensive understanding of the way in which these models prove to be superior. Finally, further research could focus on the forecast performance of the various models in bull markets and in bear markets. In the current research, only a period in which a bull markets is observed is examined. However, the forecast performance of the models could be different when the market conditions change. For example, a downward market is often accompanied with higher volatility in general.

Appendix

Table 1: Summary statistics GARCH data

	S&P 500 daily
Mean	.0004057
Median	.000586
Min	-.068954
Max	.068355
Standard deviation	.0104938
Skewness	-.3203627
Kurtosis	7.988503

Table 2: Summary statistics VIX data

	VIX
Mean	18.68297
Median	16.32
Min	9.14
Max	56.65
Standard Deviation	7.821325
Skewness	1.681939
Kurtosis	5.932182

Table 3: In-sample performance GARCH model

Intercept	0.00161
-----------	---------

GARCH Alpha	0.1137
GARCH Beta	0.86322
Jarque-Bera statistic	150.178
Kurtosis residuals	0.5648
Skewness residuals	4.0555
Box Pierce test residuals	0.6356

Table 4: In-sample performance HAR model

	Daily	Weekly	Monthly
Coefficient daily RV	0.35543	0.81199	0.26683
Coefficient weekly RV	0.23835	0.80085	0.22296
Coefficient monthly RV	0.34522	0.00565	0.93663
R-Squared	0.532	0.9566	0.9874

Table 5: Forecast encompassing regressions daily results of S&P 500. The lowest row shows which variables are significant. Is a variable is not significant, at least one of the other variables in the regression encompasses the insignificant variable. Corresponding p-values are in parentheses.

	VIX + GARCH + HAR	VIX + GARCH	VIX + HAR	GARCH + HAR
R-Squared	0.5873	0.523	0.585	0.5267
Significant	VIX (0.000) HAR (0.000)	VIX (0.000)	VIX (0.000) HAR (0.000)	HAR (0.000)
Insignificant	GARCH (0.296)	GARCH (0.063)		GARCH (0.109)

Table 6: Forecast encompassing regressions weekly results of S&P 500. The lowest row shows which variables are significant. Is a variable is not significant, at least one of the other variables in the regression encompasses the insignificant variable. Corresponding p-values are in parentheses.

	VIX + GARCH + HAR	VIX + GARCH	VIX + HAR	GARCH + HAR
R-Squared	0.6171	0.6136	0.574	0.5938
Significant	GARCH (0.038)	GARCH (0.002) VIX (0.024)	HAR (0.04)	GARCH (0.029)
Insignificant	VIX (0.119) HAR (0.606)		VIX (0.08)	HAR (0.286)

Table 7: Forecast encompassing regressions monthly results of S&P 500. The lowest row shows which variables are significant. Is a variable is not significant, at least one of the other variables in the regression encompasses the insignificant variable. Corresponding p-values are in parentheses.

	VIX + GARCH + HAR	VIX + GARCH	VIX + HAR	GARCH + HAR
R-Squared	0.8746	0.7421	0.8238	0.8746
Significant	GARCH (0.000) HAR (0.000)	VIX (0.000) GARCH (0.000)	HAR (0.000) VIX (0.002)	GARCH (0.000) HAR (0.000)
Insignificant	VIX (0.925)			

Figure 1: Plot of the GARCH data

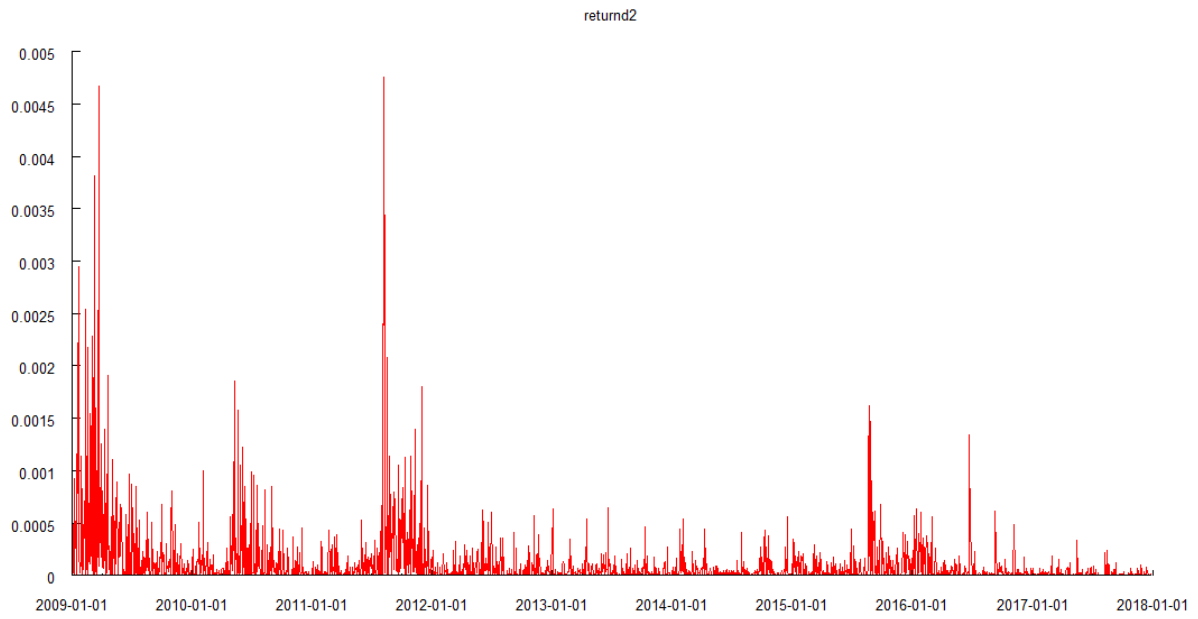


Figure 2: plot of the VIX data

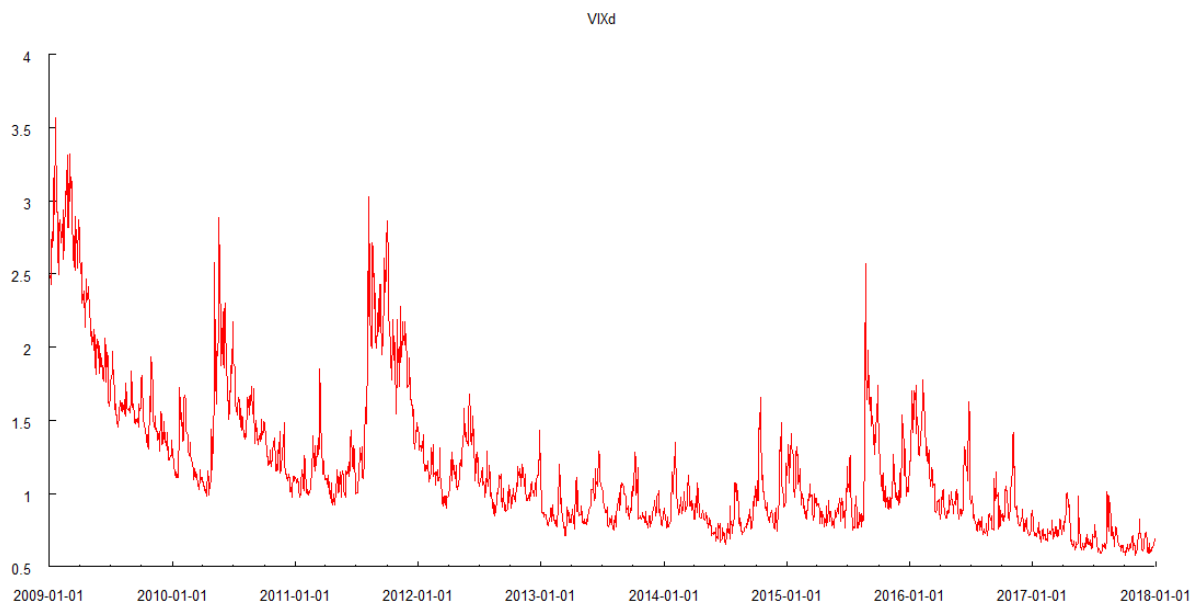


Figure 3: plot of the HAR data including the 24th of august 2015.

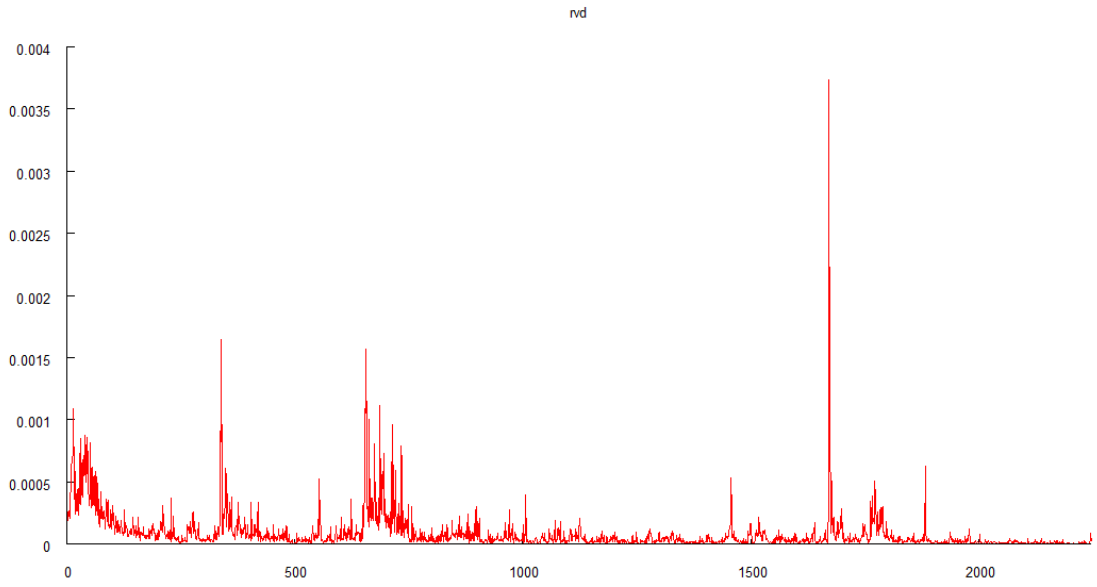
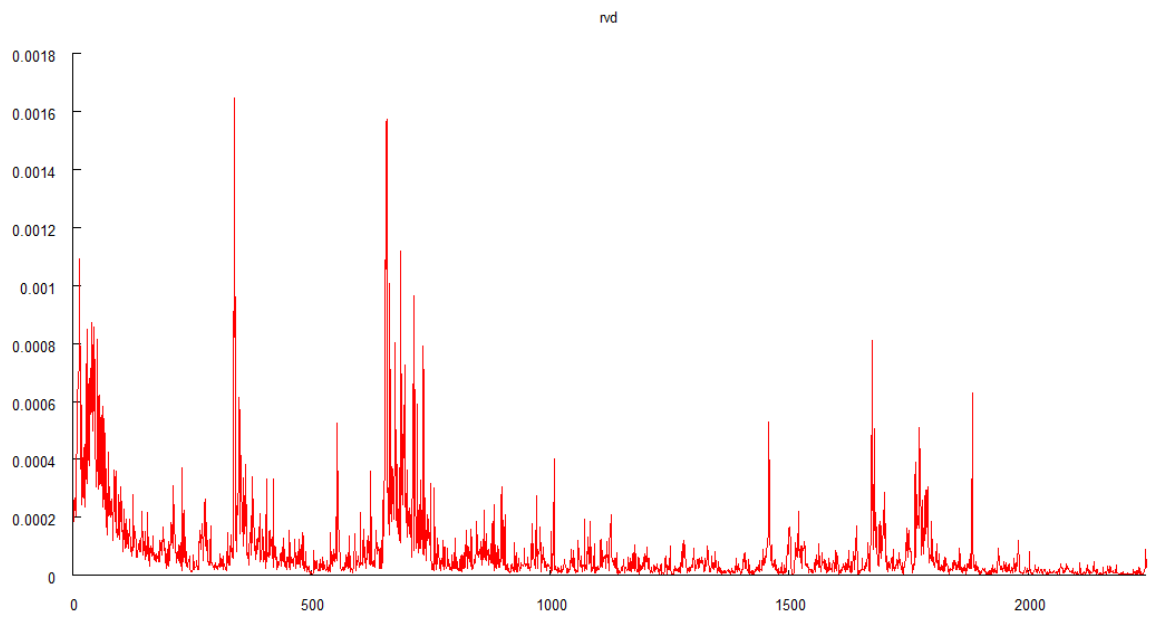


Figure 4: Har data with exclusion of the 24th of august 2015



References

- Andersen, T. G. (2000). Some reflections on analysis of high-frequency data. *Journal of Business & Economic Statistics*, 18(2), 146-153.
- Andersen, T. G., & Bollerslev, T. (1998). Answering the skeptics: Yes, standard volatility models do provide accurate forecasts. *International economic review*, 885-905.
- Andersen, T. G., Bollerslev, T., Diebold, F. X., & Labys, P. (2001). The distribution of realized exchange rate volatility. *Journal of the American statistical association*, 96(453), 42-55.
- Becker, R., Clements, A. E., & White, S. I. (2007). Does implied volatility provide any information beyond that captured in model-based volatility forecasts?. *Journal of Banking & Finance*, 31(8), 2535-2549.
- Beckers, S. (1981). Standard deviations implied in option prices as predictors of future stock price variability. *Journal of Banking & Finance*, 5(3), 363-381.
- Będowska-Sójka, B. (2015). Daily VAR forecasts with realized volatility and GARCH models. *Argumenta Oeconomica*, 34(1), 157-173.
- Black, F., & Scholes, M. (1973). The pricing of options and corporate liabilities. *Journal of political economy*, 81(3), 637-654.
- Bollen, N. P., & Whaley, R. E. (2004). Does net buying pressure affect the shape of implied volatility functions?. *The Journal of Finance*, 59(2), 711-753.
- Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of econometrics*, 31(3), 307-327.
- Breusch, T. S., & Pagan, A. R. (1979). A simple test for heteroscedasticity and random coefficient variation. *Econometrica: Journal of the Econometric Society*, 1287-1294.
- Brownlees, C., Engle, R. and B. Kelly (2012). A practical guide to volatility forecasting through calm and storm. *Journal of Risk* 14(2), 3.
- Busch, T., Christensen, B. J., & Nielsen, M. Ø. (2011). The role of implied volatility in forecasting future realized volatility and jumps in foreign exchange, stock, and bond markets. *Journal of Econometrics*, 160(1), 48-57.
- Chong, Y. Y., & Hendry, D. F. (1986). Econometric evaluation of linear macro-economic models. *The Review of Economic Studies*, 53(4), 671-690.

- Clements, M. P., & Hendry, D. F. (1993). On the limitations of comparing mean square forecast errors. *Journal of Forecasting*, 12(8), 617-637.
- Corsi, F. (2009). A simple approximate long-memory model of realized volatility. *Journal of Financial Econometrics*, 7(2), 174-196.
- Corsi, F., & Reno, R. (2009). HAR volatility modelling with heterogeneous leverage and jumps. Available at SSRN 1316953.
- Day, T. E., & Lewis, C. M. (1992). Stock market volatility and the information content of stock index options. *Journal of Econometrics*, 52(1-2), 267-287.
- Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica: Journal of the Econometric Society*, 987-1007.
- Fair, R. C., & Shiller, R. J. (1988). The informational content of ex ante forecasts.
- Fernandes, M., Medeiros, M. C., & Scharth, M. (2014). Modeling and predicting the CBOE market volatility index. *Journal of Banking & Finance*, 40, 1-10.
- Fleming, J., Ostdiek, B., & Whaley, R. E. (1993). The integration of stock, futures, and option markets: Evidence from the index derivatives. *The Journal of Futures Markets*, 16, 353-387.
- Glosten, L. R., Jagannathan, R. and D. E. Runkle (1993). On the relation between the expected value and the volatility of the nominal excess return on stocks. *Journal of Finance* 48(5), 1779– 1801.
- Griffin, J. M., & Shams, A. (2017). Manipulation in the VIX?. *The Review of Financial Studies*, 31(4), 1377-1417.
- Huang, C., Gong, X., Chen, X., & Wen, F. (2013). Measuring and forecasting volatility in Chinese stock market using HAR-CJ-M model. In *Abstract and Applied Analysis* (Vol. 2013). Hindawi.
- Huang, Z., Liu, H., & Wang, T. (2016). Modeling long memory volatility using realized measures of volatility: A realized HAR GARCH model. *Economic Modelling*, 52, 812-821.
- Kambouroudis, D. S., McMillan, D. G., & Tsakou, K. (2016). Forecasting stock return volatility: a comparison of GARCH, Implied volatility, and realized volatility models. *Journal of Futures Markets*, 36(12), 1127-1163.
- Koopman, S. J., Jungbacker, B., and Hol, E. (2005). Forecasting daily variability of the S&P100 stock index using historical, realised and implied volatility measurements. *Journal of Empirical Finance*, 12(3):445–475.

- Kourtis, A., Markellos, R. N., & Symeonidis, L. (2016). An international comparison of implied, realized, and GARCH volatility forecasts. *Journal of Futures Markets*, 36(12), 1164-1193.
- Lamoureux, C. G., & Lastrapes, W. D. (1993). Forecasting stock-return variance: Toward an understanding of stochastic implied volatilities. *The Review of Financial Studies*, 6(2), 293-326.
- Latane, H. A., & Rendleman Jr, R. J. (1976). Standard deviations of stock price ratios implied in option prices. *The Journal of Finance*, 31(2), 369-381.
- Ljung, G. M., & Box, G. E. (1978). On a measure of lack of fit in time series models. *Biometrika*, 65(2), 297-303.
- Mincer, J. A., & Zarnowitz, V. (1969). The evaluation of economic forecasts. In *Economic forecasts and expectations: Analysis of forecasting behavior and performance* (pp. 3-46). NBER.
- Sedá, P. (2012). Performance of Heterogeneous Autoregressive Models of Realized Volatility: Evidence from US Stock Market. *World Academy of Science, Engineering and Technology, International Journal of Social, Behavioral, Educational, Economic, Business and Industrial Engineering*, 6(12), 3421-3428.
- Tsay, R. S. (2000). Editor's introduction to panel discussion on analysis of high-frequency data.
- Vortelinos, D. I. (2017). Forecasting realized volatility: HAR against Principal Components Combining, neural networks and GARCH. *Research in international business and finance*, 39, 824-839.
- Whaley, R. E. (1993). Derivatives on market volatility: Hedging tools long overdue. *The journal of Derivatives*, 1(1), 71-84.
- Whaley, R. E. (2009). Understanding vix. Available at SSRN 1296743.
- White, H. (1980). A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *econometrica*, 48(4), 817-838.
- Zhang, L., Mykland, P. A., & Ait-Sahalia, Y. (2005). A tale of two time scales: Determining integrated volatility with noisy high-frequency data. *Journal of the American Statistical Association*, 100(472), 1394-1411.