Erasmus University Rotterdam

Erasmus School of Economics

Bachelor Thesis Econometrics & Operational Research

Forecasting realized volatility at an intra-day horizon using HAR defined models

Olivier van Wel (577295ow)

Supervisor: P. Vallarino

Second assessor: G. Freire

Date final version: 30 June 2024

Abstract

This paper proposes utilizing the Heterogeneous Autoregressive model of Realized Volatility (HAR-RV) model and several extensions to forecast intra-day realized volatilities and assess the ability to capture volatility characteristics over short time intervals. Empirical results reveal that by analytically selecting a system of economically accountable lag components and implementing key dummy variables, forecasting accuracy improves. The models are tested on the S&P 500 and several constituents of the Dow Jones Industrial Average. The research reports all models outperform a simplistic Autoregressive (AR) model with specifically the semivariance HAR (SHAR) model, which incorporates individual coefficients for positive and negative returns, exhibiting particularly superior results with respect to an AR model.

1 Introduction

A lot of existing research has been conducted regarding forecasting volatility, a topic of increasing importance due to the expanding availability of high-frequency data and the growth of high-frequency trading. With the rising amounts of daytraders and dense occupation on the market, the necessity for a robust and precise forecasting model has become paramount. Despite the prevalence of stochastic GARCH and heterogeneous autoregressive (HAR) models in forecasting daily realized variance, there exists a gap in the academic literature regarding their effectiveness in intra-day forecasts, such as hourly realized volatilities (RV).

This paper aims to address this gap and explores the feasibility of using simple $HAR-RV$ models, as proposed by Corsi (2009), to effectively capture volatility characteristics at an intra-day horizon and forecast at this shorter time interval. Specifically, the objective is to evaluate whether the models can improve volatility predictions compared to an AR model for one-hour ahead forecasts. Furthermore the paper examines two extensions of the Corsi (2009) paper: the HAR quarticity model ($HARQ$) as proposed by Bollerslev et al. (2015) which incorporates possible measurement errors and the semivariance HAR model ($SHAR$) defined by Patton and Sheppard (2015), which accounts for the 'leverage effect' by using $RV+$ and $RV-$. Furthermore several dummies are added to the model to incorporate anomalies including the 'weekend effect' as suggested by Kiymaz and Berument (2003), the effect of FOMC announcements and the boosted volatility of the opening hour of the market.

In the paper, solely an empirical analysis is conducted which relies on intra-day data of four different assets obtained from the TAQ database. The assets considered are the S&P 500 and three constituents of the Dow Jones Industrial Average ($DJIA$) with the data spanning from 01

January 2013 up to 30 December 2022. The raw data is tick-by-tick data that is subsequently aggregated into one-minute interval returns.

The original HAR-RV model uses lag components for the lagged day, lagged week and lagged month with economic interpretations behind these lag components. To assess which lag components would work for one-hour ahead forecasts, goodness-of-fit tests such as the Akaike information criterion (AIC) and Bayesian information criterion (BIC) are employed. Based upon these results, the best fitting models are estimated and used for predictions. Results of the various models are compared over both the estimation and prediction sample using the RMSE, QLIKE and Diebold-Mariano (DM) test as evaluation metrics.

My findings show that a model incorporating one-hour, three-hour and seven-hour lag components fits the data sets best. These specific lags embody key temporal dimensions: the one-hour reflects immediate market movements, a three-hour lag captures volatility peak persistance. The last seven-hour lag component represents the whole preceding trading day. In this paper, all models employing this lag combination outperform the corresponding AR model in prediction accuracy, suggesting an improvement in forecasting abilities. The SHAR and HARQ models perform best among the models tested and can replicate many volatility characteristics observed in the empirical data. These results contribute to the understanding and modelling of realized volatility at shorter time intervals.

The remainder of this paper is structured as follows. First, the theoretical background is provided in Section 2. Then, I elaborate on the methodological framework behind the results in Section 3 and discuss the used datasets in Section 4. Following the methodology and data, the empirical analysis is interpreted in Section 5. Lastly, the main findings are presented in Section 6.

2 Literature Review

Forecasting the volatility of assets accurately has become of utmost importance as the liquidity of stocks increases as more and more trading techniques and players join the trading market. Over the years, the choice of proxy for volatility has become the realized volatility as many studies indicate that this is a consistent unbiased estimator for the true volatility of an asset, as stated by Anderson and Bollerslev (1998) & Hansen et al. (2003). The true integrated volatility is unobservable but by taking the sum of squared intra-day returns it can be approximated it very precisely.

Müller et al. (1993) & Anderson and Bollerslev (1998) advocate that the realized variance model improves as the observation frequency increases, yielding more accurate estimates of the true volatility with a higher sample frequency. Nevertheless, the true price process remains obscured due to microstructure friction. Furthermore, Bandi and Russell (2008) shows that it is crucial to consider the trade-off between variance and the bias when estimating the variance using intra-day data hence increasing the sampling frequency increases the precision of the variance estimations but also introduces additional noise such as larger bid-ask spread fluctuations. Hansen et al. (2003) among others state that 5-min returns significantly improved reduction in noise compared to using 30-min returns, making it the conventional sample frequency in recent research.

Another significant benefit of using intra-day data was pinpointed by Bollerslev et al. (1998). They identified that intra-day data could provide a more precise assessment due to two significant features: the leverage effect and the volatility feedback effect. Taking this into account, Patton and Engle (2001) acknowledged a set of stylized facts concerning asset return volatility, which they argued should be integrated into all volatility models to provide accurate forecasts. These main characteristics are:

1. *Volatility clustering*: Return volatility fluctuates over time and a volatility shock today will impact the volatility fur multiple periods into the future.

2. *Mean reversion*: There is a level of volatility to which the volatility will eventually revert.

3. *Tail probabilities*: The unconditional distribution of asset returns exhibits heavy tails indicating a non-normal kurtosis which could range from 4 to 100.

4. *Leverage effect*: Volatility escalates more following a negative price shock compared to a equally sized positive shock.

Using straightforward component models to capturing volatility clustering and long term memory proves challenging however the HAR-RV model introduced by Corsi (2009) effectively addresses this challenge using a simple additive model. It integrates the simplistic model proposed by LeBaron (2001), which uses the sum of three distinct AR(1) processes, with the HARCH model developed by Muller et al. (1997) and Dacorogna et al. (1998) which can reproduce the long memory of volatility. The resulting model, the HAR-RV, led to a simple model that could replicate the same volatility clustering feature as well as mimicking the slowly decaying ACF of the volatility proxies. Since its introduction, this model has arguably become the preferred model for forecasting realized volatility.

The lag components of the HAR-RV model were selected with a certain economic interpretation. The lag components encompass variables for activity of the previous day, week and month. Each of these corresponds to a different trading profile: Hence the market makers who engage in high-frequency traders, the medium term traders who adjust their portfolios weekly and the institutional traders such as pension- and hedge funds who may rebalance their presumably larger portfolios on a monthly basis to curtail trading costs.

An extension of the HAR-RV model was proposed by Patton and Sheppard (2015), which incorporated the leverage effect of return volatility into the model, resulting in the semivariance HAR ($SHAR$) model. The leverage effect elucidates why negative returns tend to coincide with higher volatility. By specifically targeting downside risk within the model, Patton improved forecast accuracy compared to the HAR model developed by Corsi (2009).

Another eminent extension is the easily implementable Heterogenous Autoregressive Quarticity (HARQ) model that was developed by Bollerslev et al. (2015). It addresses the measurement errors inherent in the realized volatility due to a finite sampling frequency. The HARQ model capitalizes on the heteroskedasticity in these errors by incorporating time-varying autoregressive parameters. These parameters increase when the variance of the realized volatility is low and decrease on days with high variance. This adjustment enhances forecast persistence during stable times and facilitates quicker mean reversion during volatile periods.

The model incorporates a quarticity component that incorporates higher order moments of return, building upon the framework suggested by Corsi (2009). The quarticity, fourth moment, provides information about the distribution tails' shape and fatness. The inclusion of this additional component effectively handles the fluctuation of the magnitude in the measurement errors within the realized volatility as well as the model parameters.

Additionally it is important to consider several anomalies in return data. French (1980) first observed that due to two non-trading days preceding Monday, the stock market exhibited higher volatility and significantly lower returns than on midweek days, calling this encouter 'the weekend effect.' Most studies were done on the US stock and equity market, aligning with this research paper. This anomaly is acknowledged and dealt with using dummy variables. Besides the varying volatility between days of the week, Wood et al. (1985) states that asset returns in the opening stock exchange hour are much more volatile than other hours of the market. For this another dummy variable is created, just like the lifted volatility on days following Federal Open Market Committee (FOMC) announcements that often trigger raised volatility. These announcements inform the US on altercations of the Federal interest rate.

3 Methodological Framework

This section outlines the models and metrics used for the construction of volatility forecasts. Furthermore, it describes how the forecasts are computed and how their accuracy is assessed.

3.1 Realized Volatility

The standard continuous time process of a stock is defined as

$$dp(t) = \mu(t)dt + \sigma(t)dW(t) \tag{1}$$

where p(t) is the log of instantaneous price, $\mu(t)$ is the cadlag finite variation process, W(t) is a standard Brownian motion and $\sigma(t)$ a independent stochastic process.
For such a process, the integrated variance of the stock on day t (only looking at trading hours) is derived with

$$IV_t = \int_{t-1}^{t} \sigma^2(s)ds \tag{2}$$

To get the integrated volatility, the squareroot of equation 2 is simply taken, which coincides with the rest of the paper as volatilities are compared instead of variances.

Although the true integrated daily volatility, IV_t, is unobservable due to infinitesimal time intervals, several papers show that it can be estimated consistently using the one-day realized volatility as exhibited in a collection of papers by Andersen et al. (2001a), Andersen et al. (2001b) & Barndorff-Nielsen and Sheppard (2002). Soon thereafter, Andersen et al. (2003) showed that a simple time series model using RV outperformed popular GARCH and stochastic volatility models in forecasting which was an important contribution to forecasting literature.

The realized variance of an asset is calculated using the following equation:

$$RV_t = \sum_{i=1}^{n} r_{t,i}^2 \tag{3}$$

where n is the number of intra-day logarithmic returns and $r_{t,i}$ is the intra-day logarithmic return at time i of n within day t. To calculate the realized volatility (RV), the metric used in this paper, the squareroot of the realized variance is taken.

For all assets considered in this paper, the returns on day t at time-interval i of n, $r_{t,i}$, are calculated by:

$$r_{t,i} = log(P_{t,i}) - log(P_{t,i-1}) \tag{4}$$

Due to data limitations there is an upper bound on the number of intra-day returns, n, but by the asymptotic distribution theory of Barndorff-Nielsen and Sheppard (2002), as $n \to \infty$, we see that $\sqrt{n}(RV_t - IV_t) \xrightarrow{d} MN(0, \frac{2}{n}IQ_t)$

where $IQ_t = \int_{t-1}^{t} \sigma_s^4 ds$ stands for the Integrated Quarticity and MN stands for mixed normal.

The paper also acknowledges that high-frequency financial data can be contingent on market microstructure noise, which possibly distorts the accuracy of the realized volatility. Several methods are discussed to slightly adjust for the noise including implementing the quarticity factor into the model as explained in section 3.4.

3.2 HAR model by Corsi (2009)

The renowned heterogenous autoregressive (HAR) model is recognized for its ability in effectively capturing the main stylized features typically observed in the realized volatility despite being a simplistic AR-like model. It is widely regarded as the benchmark in volatility forecasting research as remarked in Christensen et al. (2023), among others.

The HAR model is defined as:

$$RV_t = \beta_0 + \beta_1 RV_{t-1} + \beta_2 RV_{t-1|t-5} + \beta_3 RV_{t-1|t-22} + u_t \tag{5}$$

where RV_{t-1} denotes the lagged daily RV, $RV_{t-1|t-5}$ and $RV_{t-1|t-22}$ the lagged weekly and monthly RV and u_t denotes the volatility innovation which is contemporaneously and serially independent with mean zero and a suitably truncated left tail to ensure the positivity of RV_t. $\beta_j (j = 0, 1, 2, 3)$ are unknown parameters that have to be estimated.

These are estimated using ordinary least squares (OLS) estimation. Essentially the HAR-RV model is an AR(22) model with constraints restricting coefficients in the same week (or month) to be the same.

To compute the weekly RV at day t the following formula is used which simply aggregates the RV's of the past five days:

$$RV_t^{(w)} = RV_{t-1|t-5} = \frac{1}{5}(RV_{t-1} + RV_{t-2} + RV_{t-3} + RV_{t-4} + RV_{t-5}) \tag{6}$$

Similarly the lagged monthly RV at day t is calculated as follows:

$$RV_t^{(m)} = RV_{t-1|t-22} = \frac{1}{22}(RV_{t-1} + RV_{t-2} + ... + RV_{t-21} + RV_{t-22}) \tag{7}$$

3.3 SHAR model by Patton et al. (2015)

The first variation of the HAR model that is employed in this research is semivariance HAR ($SHAR$) model defined by Patton and Sheppard (2015). This model is built to account for the 'leverage effect', a term that refers to the coherence of an asset's volatility with its returns as identified by Anderson and Bollerslev (1998). Volatility tends to stay reasonably stable when asset prices go up but when prices drop, volatility tends to increase abruptly. The SHAR model leverages these effects by using the positive and negative variance as separate parameters with unrelated coefficients.

The model is defined as follows:

$$RV_t = \beta_0 + \beta_1^- RV_{t-1}^- + \beta_1^+ RV_{t-1}^+ + \beta_2 RV_{t-1|t-5} + \beta_3 RV_{t-1|t-22} + u_t \tag{8}$$

where $RV_{t-1}^- = \sum_{i:r_{t,i}<0} r_{t,i}^2$ and $RV_{t-1}^+ = \sum_{i:r_{t,i}>0} r_{t,i}^2$ and $r_{t,i}$ is the intra-day return at time i within day t

3.4 HARQ model by Bollerslev et al. (2016)

The extension modelled by Bollerslev et al. (2015) uses an additional parameter to account for measurement errors. This parameter is known as the quarticity and refers to the fourth moment of a distribution. The true quarticity, the Integrated Quarticity (IQ), is unobservable, analogous to the IV, but can be consistently estimated using the Realized Quarticity (RQ) as exhibited by Barndorff-Nielsen and Sheppard (2002).

The Realized Quarticity is computed as follows:

$$RQ_t = \frac{n}{3}\sum_{i=1}^{n} r_{t,i}^4 \tag{9}$$

where n is the number of intra-day logarithmic returns and $r_{t,i}$ is the intra-day logarithmic return at time i of n within day t.

Including this component into the HAR-RV model gives the following:

$$\begin{aligned} RV_t = \beta_0 + \underbrace{(\beta_1 + \beta_{1Q} RQ_{t-1}^{1/2})}_{\beta_{1,t}} RV_{t-1} + \underbrace{(\beta_2 + \beta_{2Q} RQ_{t-1|t-5}^{1/2})}_{\beta_{2,t}} RV_{t-1|t-5} \\ + \underbrace{(\beta_3 + \beta_{3Q} RQ_{t-1|t-22}^{1/2})}_{\beta_{3,t}} RV_{t-1|t-22} + u_t \end{aligned} \tag{10}$$

Intuitively, the model above adjusts the HAR coefficients in proportion to the magnitude of the corresponding measurement errors. This makes far tail values of the distribution less consequential for the following forecasts. Logically the impact of the measurement errors will slowly diminish as the time goes on. This means that the adjustment is much more important for the daily lag than the weekly and monthly lagged components. Bollerslev et al. (2015) explains that adjusting solely the daily lag component is sufficient, giving us the model:

$$RV_t = \beta_0 + \underbrace{(\beta_1 + \beta_{1Q} RQ_{t-1}^{1/2})}_{\beta_{1,t}} RV_{t-1} + \beta_2 RV_{t-1|t-5} + \beta_3 RV_{t-1|t-22} + u_t \tag{11}$$

This model is referred to as the HARQ model and equation 10 is denoted as the full HARQ model (HARQ-F). This is because it permits all parameters to adjust with the estimated degree of measurement error.

3.5 Utilized model

As this paper aims to forecast one-hour ahead volatility instead of the daily RV for which the aforementioned models were designed, a few altercations are made. Below the various combinations of HAR components are illustrated along with their corresponding economic interpretation and subsequently the additional dummy variables are clarified. It is important to note that the utilized model uses subscript t as hour instead of day as done in past cited research.

The forecasting model, given the information set at time t, will look like:

$$\begin{aligned} RV_{t+1} = \beta_0 + \beta_{1,t}\, RV_{t-1|t-a} + \beta_{2,t}\, RV_{t-1|t-b} + \beta_{3,t}\, RV_{t-1|t-c} \\ +\ \beta_4\, M_t + \beta_5\, Tu_t + \beta_6\, Th_t + \beta_6\, F_t + \beta_7\, FOMC_t + \beta_8\, openingHour_t + u_t \end{aligned} \tag{12}$$

where a, b & c correspond to the lags of the models. These lag components are explained in subsection 3.5.1. The RQ can be implemented in the $\beta_{1,t}$, $\beta_{2,t}$ & $\beta_{3,t}$ depending on the model. The same applies for the SHAR model which can be switched in for the $\beta_{1,t}$. M_t, Tu_t, Th_t, F_t, $FOMC_t$ & $openingHour_t$ correspond to the dummy variables for Monday, Tuesday, Thursday, Friday, FOMC announcements (and other notable dates) and a dummy for the opening hours of the stock market. These are explained below.

3.5.1 The applied lag combinations

To apply the idea of using several lagged components of the realized volatility to forecast one-hour ahead predictions, several different lag combinations are used. The combinations are listed in Table 1 below.

Combination	**1st Lag**	**2nd Lag**	**3rd Lag**
1	1 hr	3 hrs	7 hrs
2	1 hr	3 hrs	14 hrs
3	1 hr	7 hrs	14 hrs
4	3 hr	7 hrs	14 hrs

Table 1: The Combinations of Lagged Components for the HAR(Q) models

Combinations are constructed by incorporating one short lag component, one medium lagged component, and a longer component which either encompasses a full trading day or two full trading days (13 hours). All combinations will be evaluated using the HAR-RV model, alongside the HARQ and HARQ-F models. Additionally, results for the AR(7), AR(14) and ARFIMA(5,d,0) models will be computed for comparative analysis as done in the original Corsi (2009) paper.

The 1-hour lag might provide insights into the current market movements and could be a crucial component for the model. Similarly, the 3-hour component could be interesting, as it might indicate whether volatility shocks are just beginning or have already dissipated. The 7-hour lag is selected to correspond to a complete trading day on the New York Stock Exchange (NYSE), ensuring comprehensive coverage of a day's market activity. Lastly the

14-hour component could potentially be valuable for determining if a high volatility event is lasts solely one day or extends beyond that.

3.5.2 Weekend effect

Kiymaz and Berument (2003) presents empirical evidence indicating significant difference in volatility across different days of the week. With the market opening on Monday after two days of no trading, investors react to news and announcements that occurred Friday evening or during the weekend. Friday usually tends to have higher volatility than other weekdays. Conversely, for instance Wednesday and Thursdays exhibit more stable market conditions as investors try to position themselves for the end of the trading week. To take these different days into account, dummy variables are added to the model. Adding these variables and thereby preventing potential forecasting errors resulting from these effects, is essential for the evaluation of the HAR resembling models.

3.5.3 Dummy for FOMC announcements

Several times a year the Federal Open Market Comittee (FOMC) issues statements about new monetary policies and the market condition. These announcements are interesting for traders, as these can trigger higher volatility. Calls and decisions by the Federal Reserve impact many rates including exchange and interest rates. Research by Gurkaynak et al. (2005) and later Rosa (2013) demonstrate that the volatility of US asset prices significantly rises on the day following a FOMC press release. To account for this, a dummy variable with value one is assigned to all press release dates.

Furthermore, this dummy variable also takes a value of one on dates following the result of US presidential elections. Lastly, the list includes the date of the first confirmed COVID-19 case in the US and the day COVID-19 was declared a national emergency in the US. The full list of dates is provided in the Appendix in Table 7.

3.5.4 Dummy for opening hour

Wood et al. (1985) states that on average the return of assets in the opening hour is much more volatile than other hours due to several factors. Pre-market order imbalances can cause sharp price movements and the accumulation of information outside these hours can lead to rapid price movements. To keep this in mind, an additional dummy for the first hour (9:30 - 10:00) has been implemented in the model as this hour clearly has a higher mean RV and this can oppose the bias of being a shorter time interval.

3.6 Model evaluation

To assess the forecasting accuracy of the models, three loss measures are employed: The Root Mean Squared Error (RMSE), the Mean Absolute Error (MAE) and the Quasi-Likelihood (QLIKE) function. Literature on comparison of volatility forecasts, (Patton, 2011) & (Patton and Sheppard, 2009), concludes that solely the (R)MSE and the QLIKE measures remain robust to noise in assessing forecasts and the MAE is chosen to be consistent with the foundational paper by Corsi (2009).

The first loss function, the RSME, penalizes errors on both sides symmetrically and penalizes outliers heavier than a absolute loss function would do. The RSME is specified as follows:

$$RMSE = \sqrt{\frac{\sum(RV_t - \hat{RV_t})^2}{N - P}} \tag{13}$$

where RV_i is the actual realized volatility at observation i and $\hat{RV_i}$ is the forecasted realized volatility at i. N is the number of observations and P is the number of parameter estimates.

Additionally, the MAE is calculated. The formula is given below:

$$MAE = \frac{1}{N} \sum_{t=1}^{N} |RV_t - \hat{RV_t}| \tag{14}$$

Lastly, the QLIKE loss function is incorporated. This loss function serves as a specific evaluation tool for volatility clustering, proposed by Bollerslev et al. (1994), and penalizes under-predictions heavier than over-predictions, a feature which is very appreciated in fields like risk management. The QLIKE function is a loss function that is implied by a Gaussian likelihood. The metric is defined as follows:

$$QLIKE = \frac{1}{N} \sum_{t=1}^{N} \left(\frac{RV_t}{\hat{RV_t}} - log(\frac{RV_t}{\hat{RV_t}}) - 1 \right) \tag{15}$$

These three measures are used for the out of sample evaluation. For the in-sample evaluation, solely the RMSE and MAE are measured.

Finally, to verify whether there is a significant difference in performance between the various HAR models and AR model, the one-sided Diebold-Mariano test is also used. This test is implemented in R using the dm.test from the *forecast* package. This function is founded upon the theoretical framework outlined by Harvey et al. (1997).

In addition to the Diebold-Mariano test, the Model Confidence Set (MCS) procedure is carried out to test whether the best performing model is in fact, given a level of confidence, one of the best performing models for a specific dataset. This test is implemented in R using the MCSprocedure function from the *MCS* package. This function is based on the test proposed by Hansen et al. (2011).

4 Data

This section provides a description of the various datasets used in my research and how the data has been processed.

This study strictly focusses on empirical data, examining two distinct asset types. Firstly consistent to the Corsi (2009) paper, the $S\&P\,500$ index is observed. To track this index, data of the $SPDR\ S\&P\,500\,ETF$ (SPY) has been used. This ETF is the most liquid $S\&P500$ tracker, providing a sufficient amount of intraday data. As this is a market index, it is less volatile than the second asset type that is observed. Several constituents of the Dow Jones Industrial Average ($DJIA$) as of 01 January 2013 are also observed and classified as the more volatile assets of this research. The constituents that will be examined are Coca-Cola (KO), Microsoft ($MSFT$) and American Express (AXP) each progressively increasing in volatility. Table 2 gives the summary statistics of the daily and intraday data.

Analyzing the table, it is clear that the hourly realized variances do not follow a normal distribution due to the notably high kurtosis and skewness. The datasets exhibit high positive skewness between 3.71 and 5.20, suggesting more frequent regular market volatility with occasional moments of extreme volatility, which usually coincides with sell-offs. The kurtosis values indicate that the distribution of the datasets possess fat tails. Furthermore the mean values show that the $S\&P\,500$ is logically the least volatile as it is an index of 500 constituents. It stays relatively stable if one constituent or a industry falls or spikes. Secondly, the other assets get increasingly more volatile with the consumer products industry (KO) being quite a steady market with mean (0.00280) and the finance (AXP) and technology ($MSFT$) delivering more dynamic returns.

Company	Symbol	Mean	Median	Min	Max	Kurtosis	Skewness
SPDR S&P 500 ETF	SPY	0.00240	0.00173	0.000	0.0288	27.96	3.94
Coca-Cola	KO	0.00280	0.00236	0.000	0.0470	55.40	5.20
Microsoft	MSFT	0.00367	0.00308	0.000	0.0431	27.11	3.71
American Express	AXP	0.00379	0.00304	0.000	0.0477	37.28	4.40

Table 2: Summary statistics for one-hour RVs for all discussed assets

All stock and index prices are obtained from the TAQ database and cleaned using the *highfrequency* package in *R* by Kleen, (2023). This paper focusses on prices during exchange hours only (9:30 to 16:00 EDT) from Monday to Friday. All datasets share the same sample range from 1 January 2013 up to 30 December 2022 including significant market events such as the flash crash of 24 August 2015 and the $COVID-19$ recession (2020-22). As 2020 was a very erratic year, the performance of the models is given including and excluding 2020. In contrast to the common five-minute returns, the study opts for a frequency of one-minute returns for the calculations of the realized volatility. This decision is driven by the shorter forecast horizon, targeting realized volatility over an hour instead of a day. Consequently a higher frequency per calculation of the RV is deemed crucial for accuracy and a limited bias. Furthermore, all realized volatility estimates are computed using the two scales estimator introduced by Zhang et al. (2005) to mitigate measurement errors due to noise.

The initial observations of the sample are used for model estimation, with all estimations calculated using Ordinary Leased Squares (OLS). The estimation period spans from the first observation of the various assets up to June 30, 2016, ensuring that the forecasting period is equal for all assets. Utilizing these coefficients, the remaining data is used as the forecasting sample. All estimations and forecasts are conducted using *R*.

5 Empirical analysis

In this section the empirical results are observed and analysed. In 5.1 the parameter estimates are presented and the in-sample prediction errors are given. In 5.2, the out of sample results are described.

5.1 In-sample estimation and forecast results

As briefly mentioned in Section 4, the parameters of our model in equation 12 are estimated using a simple linear regression. OLS regression estimates are consistent and normally distributed. To address potential serial correlation in the data, the Newey–West covariance correction is applied. This is done in *R* using the NeweyWest() function from the *sandwich* package. This function is founded upon the theoretical foundation addressed by Newey and West (1987).

This paper evaluates multiple combinations of different lagged RV's, making it crucial to carry out goodness-of-fit tests to determine which the optimal model for the data. Table 3 presents the information criterion measures for all models applied to the *S&P* 500 dataset. The full table of results for all datasets is provided in the appendix in Table 8 but these corroborate the findings below. Table 3 reports that the lag components of one-hour, three hour and seven hours best fit the data as the information criterion measures are the lowest for all models. This is understandable as it is important to incorporate a one-hour lag as it gives critical insights into the volatility at very short notice. Along with this lag, a combination of the three-hour lag and seven-hour lag which corresponds with a full trading day give a comprehensible interpretation of short term market behaviour. The results indicate that a 14-hour lag is redundant for estimating one-hour ahead volatility, as the volatility is predominantly affected by very short term fluctuations.

Specifically the information criterion measure are lowest for the SHAR and HARQ-F models, which could potentially imply that these models will yield the best results; this hypothesis will be checked in the subsequent subsection. As discussed above in section 3, the HAR-RV model is basically a AR(x) model which has been aggregated into three lagged components with clear economic interpretations. To check the effectiveness of this restriction, the information criterion for the unrestricted AR models, consistent to the original paper by Corsi (2009), are also given. The values for the HAR models are also superior over the AR(x) models, indicating a conceivably better fitting model in comparison to the classic AR models.

Table 4 reports the parameter estimates for the best-fitting models of Table 3 for the S&P 500 dataset. We focus the discussion on this dataset however the rest of the results and estimates are presented in Table 9 in the Appendix. The t-values are provided in parentheses below the estimated values. Below the R^2 and MSE for in-sample analysis are also given. The values of the

[1]Logically the AR(x) models remain the for all lag components but are only written once

S&P 500	AR(7)	AR(14)	HAR	SHAR	HARQ	HARQ-F
AIC						
(1,3,7)	-68860.9 [1]	-68175.2	**-68886.3**	**-69006.4**	**-68950.1**	**-69020.1**
(1,3,14)			-68700.4	-68783.8	-68751.6	-68829.6
(1,7,14)			-68808.8	-68909.2	-68867.1	-68875.9
(3,7,14)			-68344.7	-68379.5	-68342.0	-68394.2
BIC	AR(7)	AR(14)	HAR	SHAR	HARQ	HARQ-F
(1,3,7)	-68800.6	-68067.9	**-68812.5**	**-68925.9**	**-68863.0**	**-68906.1**
(1,3,14)			-68626.6	-68703.4	-68664.5	-68715.7
(1,7,14)			-68735.1	-68828.8	-68780.0	-68762.0
(3,7,14)			-68271.0	-68299.0	-68254.9	-68280.3

Table 3: Goodness-of-fit tests for the various one-hour HAR-RV models for the $S\&P500$

t-statistics indicate that the main components for the one-hour, three-hour and seven-hour lags are significant across all databases. There is one exception which is the three-hour lag for the HARQ-F model for the S&P 500 database. This could be attributed to the additional significant quarticity components added in the HARQ-F model which may create collinearity between the quarticity component and the lag component.

Furthermore it is evident that the SpecialDate dummy, which accounts for FOMC announcements, and the 9am dummy that considers the raised volatility of the opening hour of the stock exchange are also highly significant in all models. Interestingly, most weekday dummies are insignificant in most models, however the Friday is highly significant in a few cases including in all models of the S&P 500. The dummy has the lowest value of all weekdays corresponding to the potentially conservative trading behaviour before the weekend.

If we assume the lag components together give a good representation of the true volatility, we can examine the contribution of the various components to the model. The 7-hour lag, corresponding to a full trading day, appears to have the highest contribution to the realized volatility of the subsequent hour. This could be plausible as it encapsulates the activity of the market of a full day, indicating that the volatility is not solely influenced by the preceding hour. Contrarily, the three-hour lag component has a negative coefficient across all models which may suggest that raised volatility in the preceding hours, tends to subside and that periods of raised volatility do not typically extend over multiple hours in addition to seasonality in the hours of a trading day.

	HAR	SHAR	HARQ	HARQ-F
β_0	0.000240	0.000230	0.000309	0.000374
	(4.7459)	(4.6963)	(6.607)	(4.3515)
β_1	0.3809		0.372880	0.366090
	(10.8845)		(13.9533)	(12.2376)
β_2	-0.066604	0.098071	-0.090410	-0.027436
	(-1.9574)	(2.6071)	(-2.419)	(-0.6897)
β_3	0.57766	0.590410	0.583180	0.509520
	(16.2392)	(12.8586)	(15.8998)	(12.367)
$weekdayMon$	-0.000052	-0.000067	-0.000055	-0.000054
	(-1.2984)	(-1.6135)	(-1.3712)	(-1.3722)
$weekdayTue$	-0.000027	-0.000026	-0.000031	-0.000031
	(-0.7866)	(-0.7059)	(-0.8904)	(-0.8965)
$weekdayThu$	-0.000097	-0.000103	-0.000094	-0.000093
	(-2.6024)	(-2.467)	(-2.4779)	(-2.4815)
$weekdayFri$	-0.000109	-0.000129	-0.000113	-0.000113
	(-3.2326)	(-3.473)	(-3.3091)	(-3.3603)
$SpecialDate$	0.000328	0.000372	0.000323	0.000326
	(3.3384)	(3.295)	(3.2704)	(3.316)
$dummy\,9am$	0.000108	0.000195	0.000120	0.000096
	(3.1920)	(5.4671)	(3.3169)	(2.5759)
β_1 +		0.050816		
		(2.4423)		
β_1 -		0.187100		
		(12.4971)		
β_{1Q}			1.223700	1.718300
			(2.3848)	(2.748)
β_{2Q}				-3.501400
				(-5.1787)
β_{3Q}				5.090900
				(2.1034)

Table 4: Ordinary least squares Model Estimates for S&P 500

Lastly, it is noteworthy to mention that when looking at all datasets, as the asset volatility increases, the coefficient of the seven-hour lag reduces, rendering it less of an influence on the forecasted realized volatility. In addition to that, comparing the coefficients over the various assets reveals that the β_1- coefficient in the SHAR model, which represents negative returns, declines as asset volatility increases and the positive returns coefficient , β_1+, increases. This can be explained by considering that for individual stocks, company announcements have a big impact on the stock price in both directions and for the index asset, an individual company's announcement will not have much effect on the price, whereas negative industry- or nation-wide announcements will.

	S&P500		**KO**		**MSFT**		**AXP**	
	RMSE	**MAE**	**RMSE**	**MAE**	**RMSE**	**MAE**	**RMSE**	**MAE**
AR(7)	0.000791	0.000507	0.001421	0.000821	0.001486	0.001033	0.001446	0.000914
HAR	0.000789	0.000509	0.001050	0.000720	0.001458	0.000984	0.001384	0.000845
SHAR	**0.000781**	**0.000499**	**0.001026**	**0.000700**	**0.001448**	**0.000976**	**0.001375**	**0.000828**
HARQ	0.000790	0.000514	0.001051	0.000723	0.001458	0.000984	0.001384	0.000850
HARQ-F	0.000792	0.000521	0.001051	0.000722	0.001459	0.000985	0.001385	0.000856

Table 5: One-hour ahead in-sample performance

Table 5 reports the in-sample performance of the evaluated models. Consistent with the methodology proposed by Corsi (2009), the performance of the models is evaluated using the root mean square error (RMSE) and the mean absolute error (MAE). The results clearly indicate that HAR models perform better than the conventional AR(7) model as the MAE and RSME are lower across all models. Notably the SHAR demonstrates the most pronounced improvement, which may be explainable due to its ability to utilize the correlation between negative returns and high volatility.

5.2 Out of sample forecasts results

The out of sample results presented in this section exclude the year 2020, due to the erratic market conditions. These results are presented in Table 6, while the results including the year 2020 are given in the Appendix in Table 10. Across all datasets, the HAR models outperform the AR(7) model as the RMSE and QLIKE are lower for all datasets except the RMSE of the S&P 500. Notably, the SHAR model exhibits the best results, as highlighted in bold, and the Quarticity models perform very similarly to the original HAR model by Corsi (2009).

To ascertain whether the SHAR model significantly differs from the AR(7) model, the Diebold-Mariano (DM) test is applied. The hypothesis being tested here is $H_0 : MSE_{SHAR} = MSE_{AR(7)}$ against a one-sided alternative $H_\alpha : MSE_{SHAR} < MSE_{AR(7)}$. As indicated in the table, null hypotheses are rejected for the KO and MSFT datasets, implying that the SHAR significantly outperforms at a significance level of 0.003% and 7% respectively. This suggests that the SHAR model is more effective for individual stocks than for the index, potentially due to the stronger leverage effect for individual stocks.

In addition to the DM test, the MSC procedure is employed using all computed models to check the superior set of models. These results are provided in Table 11. Interestingly, the superior sets all include either the HARQF or HARQ model but solely the MSFT set includes our overal best performing model, the SHAR model. Conforming with our previous results, all sets do not include the AR(7) model. Future research could explore the performance of an ARFIMA in comparison to the HAR models to determine whether it outperforms these models or compliments them in the superior set.

	S&P500			**KO**			**MSFT**			**AXP**		
Excluding 2020	**RMSE**	**MAE**	**QLIKE**	**RMSE**	**MAE**	**QLIKE**	**RMSE**	**MAE**	**QLIKE**	**RMSE**	**MAE**	**QLIKE**
AR(7)	0.000974	0.000606	0.351075	0.001045	0.000717	0.284935	0.001397	0.000956	0.317154	0.001800	0.001072	0.387675
HAR	0.000977	0.000611	0.345209	0.001057	0.000691	0.249309	0.001355	0.000921	0.279233	0.001736	0.001028	0.338677
SHAR	**0.000971**	**0.000601**	**0.330946**	**0.000992**	**0.000679**	**0.238402**	**0.001341**	**0.000911**	**0.275513**	**0.001718**	**0.001010**	**0.324997**
HARQ	0.000979	0.000618	0.341274	0.001009	0.000697	0.250088	0.001359	0.000925	0.280286	0.001736	0.001033	0.336638
HARQ-F	0.000979	0.000625	0.339163	0.001010	0.000697	0.250493	0.001359	0.000925	0.279970	0.001738	0.001038	0.334450
	DM test with AR(7)			**DM test with AR(7)**			**DM test with AR(7)**			**DM test with AR(7)**		
	DM t-statistic	p-value		DM t-statistic	p-value		DM t-statistic	p-value		DM t-statistic	p-value	
SHAR	-0.49	0.31		-7.10	0.003		-1.44	0.07		-0.70	0.24	

Table 6: One-hour ahead out of sample performance of the models excluding 2020

Table 6 also reveals that plausibly all models perform better on less volatile assets with generally lower loss metrics, as these are easier to forecast due to the smaller fluctuations. However, contrarily the QLIKE results are not necessarily better for the less volatile assets such as the S&P 500 and KO. This discrepancy may arise from the fact that the QLIKE loss function penalizes under predictions more severely than over predictions and possibly the fitted models for the more stable assets do not work as well when volatility spikes occur.

Figures 5, 6, 7 and 8 in the Appendix show the various forecasts for a random sample spanning from November to December 2016. They illustrate the differences among forecasts between the AR(7), HAR, HARQ-F and SHAR models. Particularly in Figure 6, it is visible that the HAR models exhibit more fluctuations in volatility forecasts compared to the AR(7) model and follow the preceding observation more. This reaffirms our earlier conclusion that the addition of various dummies and extensions, like incorporating quarticity of semivariance, helps the model to capture the dynamic characteristic of hourly RV better.

Given these findings, a potential area for future research would be modifying the model into a state-space model, as proposed by Durbin and Koopman (2012). This modification could potentially enhance the model's ability in capturing seasonality more effectively than the current approach with dummy variables. Figures 1, 2, 3 and 4 show the Partial Autocorrelation Function of the datasets which all show signs of daily seasonality as there are spikes every seven observations. This could further improve the adaptability of the model to different market conditions and capture the volatility clustering effect better.

6 Conclusion

This research investigates whether the renowned HAR-RV model proposed by Corsi (2009) would work at a short forecasting interval and could capture volatility characteristics at an intra-day horizon. Empirical analysis on multiple datasets consisting of stable indices and volatile stocks shows that with specific interpretable lags, the HAR model and various extensions capture the characteristics at this interval more effectively than the simple AR model. The selection of lagged components that best fit the data are lags of the preceding hour, three hours and the last full trading day consisting of the last seven hours. These lags can be interpreted as the last hour which shows what is happening at a very short notice, a three hour lag indicating whether the possible volatility spike is short lasting or of longer duration and a seven hour lag with the highest coefficient in all model which informs the model on the general level of volatility over the past day.

The HAR model and its extensions, the SHAR, HARQ and HARQ-F, yield more accurate forecasts of one-hour ahead realized volatility than the AR model. The SHAR model steadily outperforms all models and following a Diebold-Mariano test significantly outperforms the AR model. This is potentially due to the captured leverage effect which was first incorporated into the model by Patton and Sheppard (2015). This effect is influential at short interval as it coincides with raised volatility.

Further research could investigate whether this model can perform at the same level as other stochastic volatility models such as GARCH, ARFIMA and Mixed Data Sampling (MIDAS) models that are often used for high frequency data. Incorporating the state-space system of Durbin and Koopman (2012) into the model could be interesting as well, as it will likely mitigate daily seasonal effects in the data. A final proposal for further research is utilizing a moving window for the estimation of the coefficients as this could improve performance too.

References

T. Andersen, T. Bollerslev, F. Diebold, and P. Labys. The distribution of stock returns' volatility. *Journal of Financial Economics 61 43-76*, 2001a.

T. Andersen, T. Bollerslev, F. Diebold, and H. Ebens. The distribution of realized exchange rate volatility. *Journal of the American Statistical Association 96 42-65*, 2001b.

T. Andersen, T. Bollerslev, F. Diebold, and P. Labys. Modeling and forecasting realized volatility. *Econometrica 71: 579-625*, 2003.

T. Anderson and T. Bollerslev. Answering the skeptics: Yes, standard volatility models do provide accurate forecasts. *International Economic Review 39(4) 885-905*, 1998.

F. Bandi and J. Russell. Microstructure noise, realized variance and optimal sampling. *The review of Economic Studies 75(2) 339-369*, 2008.

O. Barndorff-Nielsen and K. Sheppard. Econometric analysis of realized volatility and its use in estimating stochastic volatility models. *Journal of the Royal Statistical Society 64 253-280*, 2002.

T. Bollerslev, R. Engle, and D. Nelson. Chapter 49 arch models. *Handbook of Econometrics (pp. 2959-3038)*, 1994.

T. Bollerslev, J. Litvinova, and G. Tauchen. Leverage and volatility feedback effects in high-frequency data. *Journal of Financial Econometrics 4(3) 353-384*, 1998.

T. Bollerslev, A. Patton, and R. Quaedvlieg. Exploiting the errors: A simple approach for improved volatility forecasting. *Journal of Econometrics 192 1-18*, 2015.

K. Christensen, M. Siggaard, and B. Veliyev. A machine learning approach to volatility forecasting. *Journal of Financial Econometrics 21(5) 1680-1727*, 2023.

F. Corsi. A simple approximate long-memory model of realized volatility. *Journal of Financial Econometrics 7(2) 174-196*, 2009.

M. Dacorogna, R. Dav, R. Olsen, O. Pictet, and U. Muller. Modelling short-term volatility with garch and harch models. *Nonlinear Modelling of High Frequency Financial Time Series 161-176*, 1998.

J. Durbin and S. Koopman. Time series analysis by state space methods. *Oxford Statistical Science Series*, 2012.

K. French. Stock returns and the weekend effect. *Journal of Financial Economics 8(1) 55-69*, 1980.

R. Gurkaynak, B. Sack, and E. Swanson. Do actions speak louder than words? the response of asset prices to monetary policy actions and statements. *International Journal of Central Banking 1 55-93*, 2005.

P. Hansen, A. Lunde, and J. Nason. Choosing the best volatility models: The model confidence set approach. *Oxford Bulletin of Economics and Statistics 65(1) 839-861*, 2003.

P. Hansen, A. Lunde, and J. Nason. The model confidence set. *Econometrica, 79(2), 453-497.*, 2011.

D. Harvey, S. Leybourne, and P. Newbold. Testing the equality of prediction mean squared errors. *International Journal of Forecasting 13(2) 281-291*, 1997.

H. Kiymaz and H. Berument. The day of the week effect on stock market volatility and volume: International evidence. *Review of Financial Economics 12 363-380*, 2003.

B. LeBaron. Stochastic volatility as a simple generator of financial power laws and long memory. *Quantitative Finance 1 621-631*, 2001.

U. Muller, U. Dacorogna, R. Dav, R. Olsen, O. Pictet, and J. von Weizsacker. Volatilities of diffrent time resolutions - analysing the dynamics of market components. *Journal of Empirical Finance 4 213-239*, 1997.

U. Müller, M. Dacorogna, R. Davé, O. Pictet, R. Olsen, and J. Ward. Fractals and intrinsic time - a challenge to econometricians. *University of Zurich, Olsen Associates*, 1993.

W. Newey and K. West. A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica 55(3), 703-708*, 1987.

A. Patton. Volatility forecast comparison using imperfect volatility proxies. *Journal of Econometrics 160: 246-256*, 2011.

A. Patton and R. Engle. What good is a volatility model? *Quantitative Finance 1(1) 237-245*, 2001.

A. Patton and K. Sheppard. Evaluating volatility and correlation forecasts. *Handbook of Financial Time Series: 801-839*, 2009.

A. Patton and K. Sheppard. Good volatility, bad volatility: Signed jumps and the persistance of volatility. *Review of Economics and Statistics 97(3) 683-697*, 2015.

C. Rosa. The financial market effect of fomc minutes. *Economic Policy Review 1 67-81*, 2013.

R. Wood, T. McInish, and J. Ord. The behavior of prices on wall street: The effects of information and of traders. *Journal of Finance 40(3), 723-739*, 1985.

L. Zhang, Y. A ıt Sahalia, and P. Mykland. A tale of two time scales: Determining integrated volatility with noisy high frequency data. *Journal of the American Statistical Association 100, 1394–1411*, 2005.

7 Appendix

FOMC Announcement Dates 2013 - 2022

2013	**2016**	**2019**	**2022**
30 January 2013	27 January 2016	30 January 2019	26 January 2022
20 March 2013	16 March 2016	20 March 2019	16 March 2022
1 May 2013	27 April 2016	1 May 2019	4 May 2022
19 June 2013	15 June 2016	19 June 2019	15 June 2022
31 July 2013	27 July 2016	31 July 2019	21 September 2022
18 September 2013	21 September 2016	18 September 2019	2 November 2022
16 October 2013	2 November 2016	4 October 2019	14 December 2022
30 October 2013	14 December 2016	30 October 2019	
18 December 2013		11 December 2019	
2014	**2017**	**2020**	**Other**
29 January 2014	1 February 2017	29 January 2020	*Presidential Elections:*
4 March 2014	15 March 2017	3 March 2020	3 November 2020
19 March 2014	3 May 2017	15 March 2020	8 November 2016
30 April 2014	14 June 2017	23 March 2020	*US Covid Announcements:*
18 June 2014	26 July 2017	31 March 2020	1 March 2020
30 July 2014	20 September 2017	29 April 2020	13 March 2020
17 Setember 2014	1 November 2017	10 June 2020	
29 October 2014	13 December 2017	29 July 2020	
17 December 2014		27 August 2020	
		16 September 2020	
		5 November 2020	
		16 December 2020	
2015	**2018**	**2021**	
28 January 2015	31 January 2018	27 January 2021	
18 March 2015	21 March 2018	17 March 2021	
29 April 2015	2 May 2018	28 April 2021	
17 June 2015	13 June 2018	16 June 2021	
29 July 2015	1 August 2018	28 July 2021	
17 September 2015	26 September 2018	22 September 2021	
28 October 2015	8 November 2018	3 November 2021	
16 December 2015	19 December 2018	15 December 2021	

Table 7: FOMC Announcement Dates

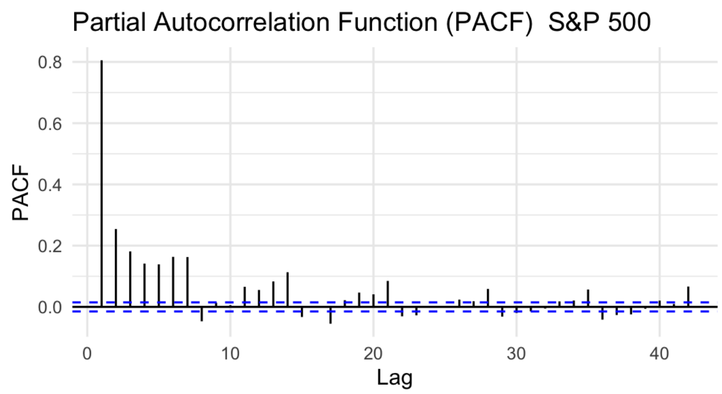


Figure 1: PACF for the S&P500

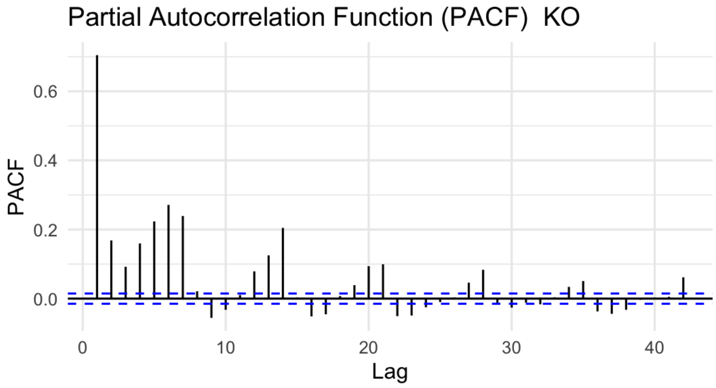


Figure 2: PACF for KO

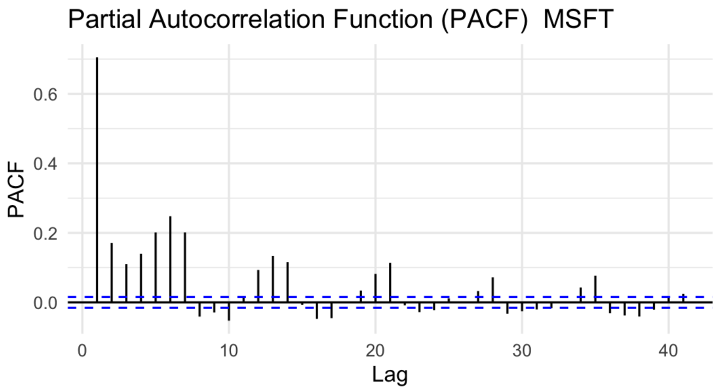


Figure 3: PACF for MSFT

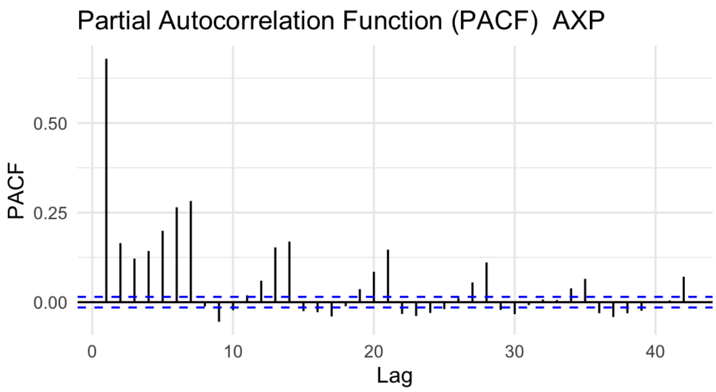


Figure 4: PACF for AXP

S&P 500	**AR(7)**	**AR(14)**	**HAR**	**SHAR**	**HARQ**	**HARQ-F**
AIC						
(1,3,7)	-68860.9	-68175.2	**-68886.3**	**-69006.4**	**-68950.1**	**-69020.1**
(1,3,14)			-68700.4	-68783.8	-68751.6	-68829.6
(1,7,14)			-68808.8	-68909.2	-68867.1	-68875.9
(3,7,14)			-68344.7	-68379.5	-68342.0	-68394.2
BIC						
(1,3,7)	-68800.6	-68067.9	**-68812.5**	**-68925.9**	**-68863.0**	**-68906.1**
(1,3,14)			-68626.6	-68703.4	-68664.5	-68715.7
(1,7,14)			-68735.1	-68828.8	-68780.0	-68762.0
(3,7,14)			-68271.0	-68299.0	-68254.9	-68280.3

KO	**AR(7)**	**AR(14)**	**HAR**	**SHAR**	**HARQ**	**HARQ-F**
AIC						
(1,3,7)	-64684.8	-64916.4	**-65721.4**	**-65448.3**	**-65538.3**	**-65532.6**
(1,3,14)			-65594.4	-65346.2	-65434.3	-65429.4
(1,7,14)			-65665.7	-65378.5	-65473.9	-65480.9
(3,7,14)			-64988.4	-64822.0	-64911.4	-64926.2
BIC						
(1,3,7)	-64624.5	-64809.2	**-65640.9**	**-65374.6**	**-65451.1**	**-65418.6**
(1,3,14)			-65513.9	-65272.5	-65347.1	-65315.4
(1,7,14)			-65585.3	-65304.8	-65386.8	-65366.9
(3,7,14)			-64907.9	-64748.2	-64824.2	-64812.2

AXP	**AR(7)**	**AR(14)**	**HAR**	**SHAR**	**HARQ**	**HARQ-F**
AIC						
(1,3,7)	-61395.8	-61572.2	**-62202.1**	**-62124.9**	**-62153.5**	**-62159.4**
(1,3,14)			-62088.5	-62020.5	-62047.9	-62053.6
(1,7,14)			-62119.3	-62064.6	-62093.5	-62110.6
(3,7,14)			-61578.9	-61563.2	-61621.8	-61631.2
BIC						
(1,3,7)	-61335.5	-61465.0	**-62121.7**	**-62051.1**	**-62066.4**	**-62045.4**
(1,3,14)			-62008.1	-61946.8	-61960.8	-61939.7
(1,7,14)			-62038.8	-61990.9	-62006.4	-61996.7
(3,7,14)			-61498.5	-61489.4	-61534.6	-61517.3

MSFT	**AR(7)**	**AR(14)**	**HAR**	**SHAR**	**HARQ**	**HARQ-F**
AIC						
(1,3,7)	-60956.2	-61110.9	**-61561.3**	**-61484.5**	**-61530.6**	**-61530.9**
(1,3,14)			-61407.0	-61340.2	-61388.6	-61391.8
(1,7,14)			-61473.0	-61409.4	-61459.7	-61463.4
(3,7,14)			-60830.9	-60837.2	-60892.6	-60901.1
BIC						
(1,3,7)	-60787.2	-61003.7	**-61480.9**	**-61410.8**	**-61443.5**	**-61416.9**
(1,3,14)			-61326.6	-61266.5	-61301.5	-61277.9
(1,7,14)			-61392.6	-61335.7	-61372.6	-61349.5
(3,7,14)			-60750.5	-60763.5	-60805.4	-60787.2

Table 8: Goodness-of-fit test for the various one-hour HAR-RV models for all four datasets (S&P 500, KO, AXP & MSFT)

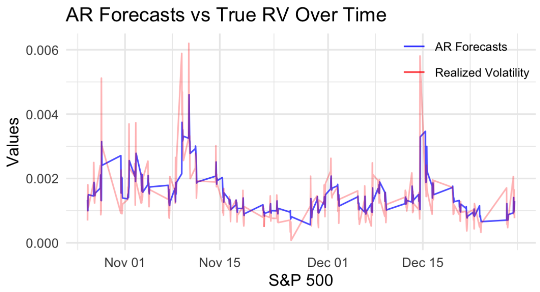
	S&P500				KO					MSFT				AXP			
	HAR	SHAR	HARQ	HARQ-F	HAR	SHAR	HARQ	HARQ-F		HAR	SHAR	HARQ	HARQ-F	HAR	SHAR	HARQ	HARQ-F
β_0	0.000240	0.000230	0.000309	0.000374	0.000388	0.000340	0.000450	0.000468	β_0	0.000501	0.000474	0.000545	0.000550	0.000529	0.000504	0.000584	0.000675
	(4.7459)	(4.6963)	(6.607)	(4.3515)	(5.9815)	(5.7757)	(6.0794)	(6.1162)		(5.936)	(5.2059)	(5.8905)	(6.0269)	(5.5587)	(5.8484)	(6.0795)	(5.4884)
β_1	0.380900		0.372880	0.366090	0.363990		0.323090	0.323130	β_1	0.385200		0.359990	0.363430	0.351110		0.336110	0.341960
	(10.8845)		(13.9533)	(12.2376)	(17.3499)		(16.4115)	(16.5026)		(15.717)		(17.0836)	(18.0775)	(22.9997)		(17.5358)	(17.9875)
β_2	-0.066604	0.098071	-0.090410	-0.027436	-0.138140	-0.003364	-0.123190	-0.126330	β_2	-0.108940	0.097306	-0.098236	-0.119750	-0.069120	0.090879	-0.063546	-0.090127
	(-1.9574)	(2.6071)	(-2.419)	(-0.6897)	(-5.7347)	(-0.1263)	(-4.8467)	(-4.9641)		(-4.117)5	(2.8127)	(-3.3952)	(-4.2856)	(-2.6514)	(3.2469)	(-2.1916)	(-3.0101)
β_3	0.577660	0.590410	0.583180	0.509520	0.550310	0.554640	0.556210	0.551590	β_3	0.530290	0.509320	0.533400	0.551040	0.477060	0.435510	0.476290	0.477630
	(16.2392)	(12.8586)	(15.8998)	(12.367)	(17.6923)	(15.7353)	(17.6603)	(18.1528)		(16.8609)	(14.3698)	(17.2346)	(18.0022)	(13.2171)	(10.4532)	(12.5824)	(11.017)
weekdayMon	-0.000052	-0.000067	-0.000055	-0.000054	-0.000040	-0.000038	-0.000040	-0.000042	weekdayMon	-0.000104	-0.000110	-0.000108	-0.000108	-0.000126	-0.000122	-0.000122	-0.000121
	(-1.2984)	(-1.6135)	(-1.3712)	(-1.3722)	(-0.8701)	(-0.8552)	(-0.9073)	(-0.9791)		(-1.4984)	(-1.5318)	(-1.7199)	(-1.7395)	(-2.1806)	(-2.0188)	(-2.1383)	(-2.1362)
weekdayTue	-0.000027	-0.000026	-0.000031	-0.000031	0.000046	0.000042	0.000034	0.000032	weekdayTue	0.000001	-0.000007	-0.000004	-0.000004	-0.000061	-0.000051	-0.000059	-0.000058
	(-0.7866)	(-0.7059)	(-0.8904)	(-0.8965)	(1.1542)	(1.0205)	(0.8343)	(0.7652)		(0.0103)	(-0.1213)	(-0.0728)	(-0.0717)	(-1.3286)	(-1.0122)	(-1.285)	(-1.2776)
weekdayThu	-0.000097	-0.000103	-0.000094	-0.000093	-0.000051	-0.000047	-0.000050	-0.000049	weekdayThu	-0.000083	-0.000103	-0.000084	-0.000085	-0.000021	-0.000019	-0.000026	-0.000031
	(-2.6024)	(-2.467)	(-2.4779)	(-2.4815)	(-1.3771)	(-1.1895)	(-1.2685)	(-1.2196)		(-1.4795)	(-1.6516)	(-1.4522)	(-1.4574)	(-0.4055)	(-0.3421)	(-0.4949)	(-0.5882)
weekdayFri	-0.000109	-0.000129	-0.000113	-0.000113	-0.000052	-0.000048	-0.000052	-0.000051	weekdayFri	-0.000011	-0.000075	-0.000029	-0.000036	-0.000008	-0.000003	-0.000008	-0.000012
	(-3.2326)	(-3.473)	(-3.3091)	(-3.3603)	(-1.4583)	(-1.2539)	(-1.3386)	(-1.3169)		(-0.1884)	(-1.2764)	(-0.4978)	(-0.6189)	(-0.1292)	(-0.0416)	(-0.1361)	(-0.1983)
SpecialDate	0.000328	0.000372	0.000323	0.000326	0.000300	0.000328	0.000308	0.000309	SpecialDate	0.000267	0.000289	0.000272	0.000272	0.000403	0.000430	0.000409	0.000414
	(3.3384)	(3.295)	(3.2704)	(3.316)	(3.6683)	(3.5654)	(3.3348)	(3.3366)		(2.6486)	(2.4847)	(2.4318)	(2.4252)	(3.9282)	(3.7288)	(3.7365)	(3.7682)
dummy 9am	0.000108	0.000195	0.000120	0.000096	0.001389	0.001460	0.001400	0.001399	dummy 9am	0.001554	0.001691	0.001563	0.001556	0.001680	0.001766	0.001689	0.001680
	(3.1920)	(5.4671)	(3.3169)	(2.5759)	(21.0194)	(22.8199)	(22.0008)	(21.9377)		(17.5451)	(19.4615)	(18.1113)	(18.0689)	(20.0328)	(22.0004)	(20.3874)	(20.5018)
β_1+		0.050816				0.115280			β_1+		0.105050				0.157570		
		(2.4423)				(5.8227)					(5.5071)				(8.8606)		
β_1-		0.187100				0.165550			β_1-		0.134170				0.122930		
		(12.4971)				(9.1742)					(4.5177)				(7.2671)		
β_{1Q}			1.223700	1.718300			0.863280	0.881340	β_{1Q}			0.130000	0.117120			0.550070	0.408930
			(2.3848)	(2.748)			(1.6214)	(1.714)				(0.4969)	(0.4431)			(2.8876)	(2.1813)
β_{2Q}				-3.501400				0.063090	β_{2Q}				-0.061858				0.155840
				(-5.1787)				(0.1338)					(-0.3818)				(0.2196)
β_{3Q}				5.090900				-0.379650	β_{3Q}				0.247100				0.721040
				(2.1034)				(-0.4347)					(0.5726)				(1.4176)

Table 9: Model estimates for all datasets (SP 500, KO, MSFT AMX)

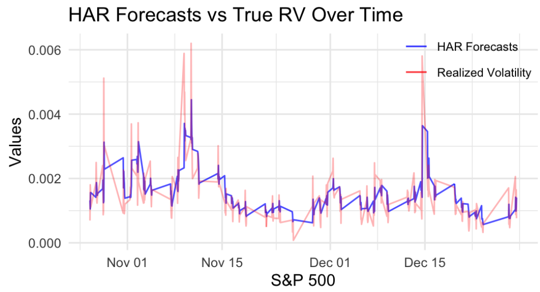
	S&P500			KO			MSFT			AXP		
Excluding 2020	**RMSE**	**MAE**	**QLIKE**	**RMSE**	**MAE**	**QLIKE**	**RMSE**	**MAE**	**QLIKE**	**RMSE**	**MAE**	**QLIKE**
AR(7)	0.000974	0.000606	0.351075	0.001045	0.000717	0.284935	0.001397	0.000956	0.317154	0.001800	0.001072	0.387675
HAR	0.000977	0.000611	0.345209	0.001057	0.000691	0.249309	0.001355	0.000921	0.279233	0.001736	0.001028	0.338677
SHAR	**0.000971**	**0.000601**	**0.330946**	**0.000992**	**0.000679**	**0.238402**	**0.001341**	**0.000911**	**0.275513**	**0.001718**	**0.001010**	**0.324997**
HARQ	0.000979	0.000618	0.341274	0.001009	0.000697	0.250088	0.001359	0.000925	0.280286	0.001736	0.001033	0.336638
HARQ-F	0.000979	0.000625	0.339163	0.001010	0.000697	0.250493	0.001359	0.000925	0.279970	0.001738	0.001038	0.334450

	S&P500			KO			MSFT			AXP		
Including 2020	**RMSE**	**MAE**	**QLIKE**	**RMSE**	**MAE**	**QLIKE**	**RMSE**	**MAE**	**QLIKE**	**RMSE**	**MAE**	**QLIKE**
AR(7)	**0.001109**	**0.000667**	0.353016	0.001353	0.000814	0.295793	0.001689	0.001068	0.320059	0.002118	0.001205	0.380537
HAR	0.001112	0.000674	0.348700	0.001311	0.000790	0.260872	**0.001166**	0.001037	0.285942	0.002051	0.001162	0.334661
SHAR	0.001136	0.000671	**0.337241**	**0.001300**	**0.000774**	**0.248266**	0.001658	**0.001030**	**0.283390**	**0.002046**	**0.001142**	**0.321428**
HARQ	0.001118	0.000680	0.344878	0.001322	0.000797	0.262101	0.001662	0.001042	0.287156	0.002054	0.001167	0.332964
HARQ-F	0.001119	0.000687	0.342508	0.001326	0.000798	0.262919	0.001662	0.001043	0.286792	0.002063	0.001173	0.331577

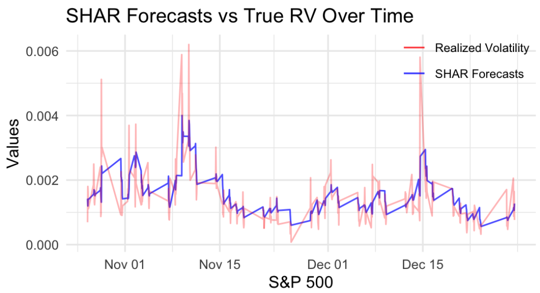
Table 10: One-hour ahead out-sample performance



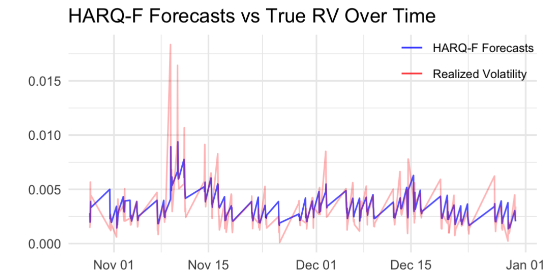
(a) AR vs Realized Volatility



(b) HAR vs Realized Volatility

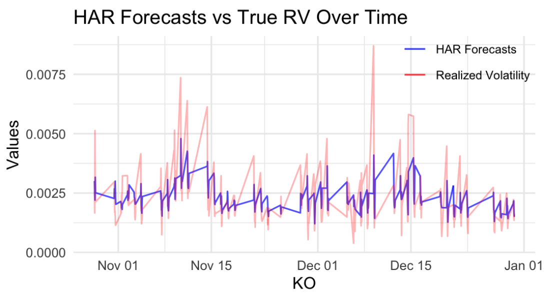


(c) SHAR vs Realized Volatility

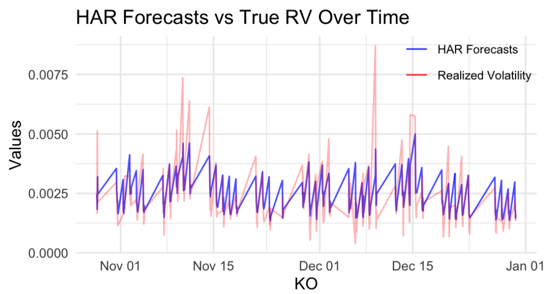


(d) HARQ-F vs Realized Volatility

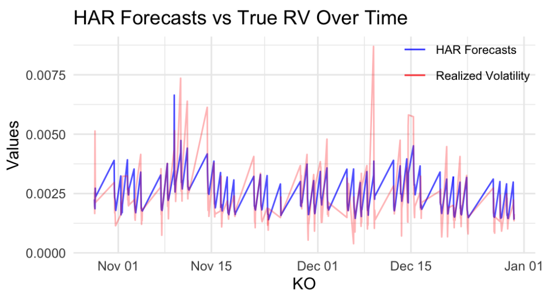
Figure 5: Realized volatility compared with the computed HAR models for S&P500, Nov 2016 - Dec 2016



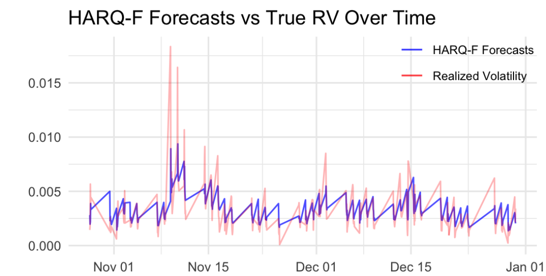
(a) AR vs Realized Volatility



(b) HAR vs Realized Volatility

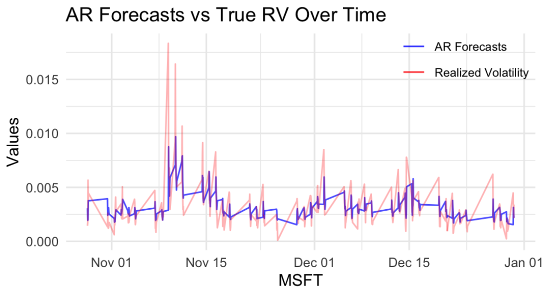


(c) SHAR vs Realized Volatility

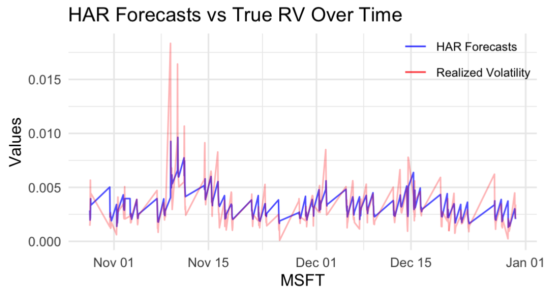


(d) HARQ-F vs Realized Volatility

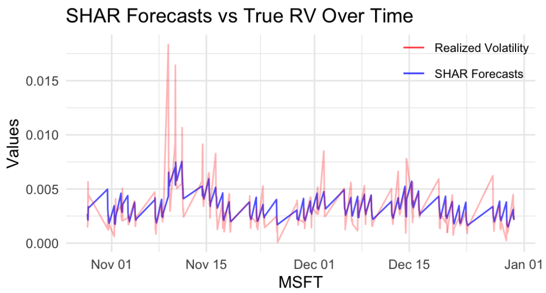
Figure 6: Realized volatility compared with the computed HAR models for KO, Nov 2016 - Dec 2016



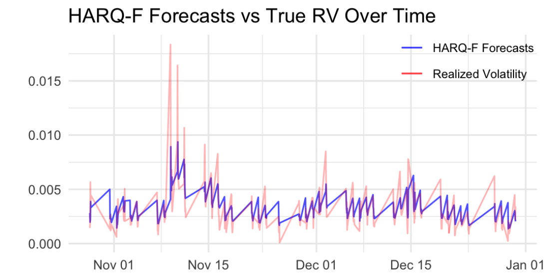
(a) AR vs Realized Volatility



(b) HAR vs Realized Volatility

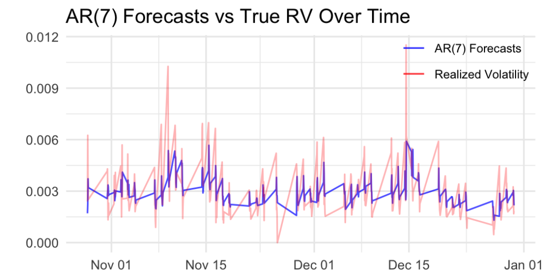


(c) SHAR vs Realized Volatility

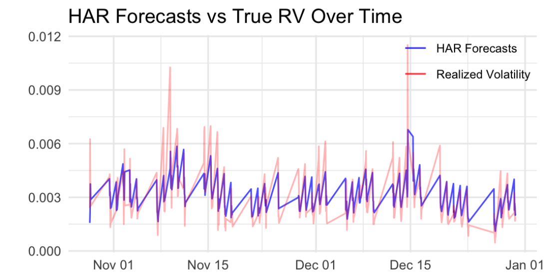


(d) HARQ-F vs Realized Volatility

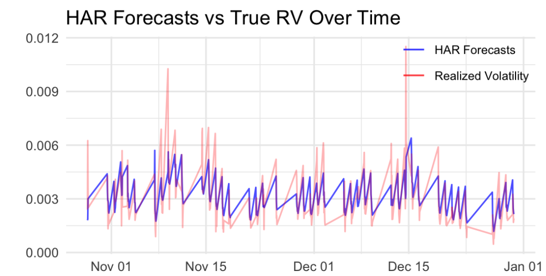
Figure 7: Realized volatility compared with the computed HAR models for MSFT, Nov 2016 - Dec 2016



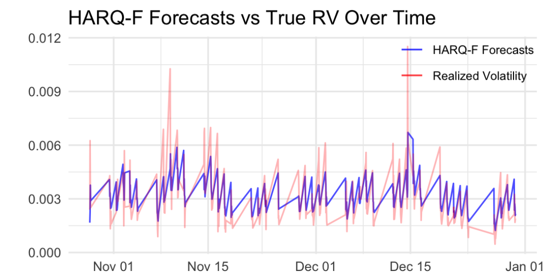
(a) AR vs Realized Volatility



(b) HAR vs Realized Volatility



(c) SHAR vs Realized Volatility



(d) HARQ-F vs Realized Volatility

Figure 8: Realized volatility compared with the computed HAR models for AXP, Nov 2016 - Dec 2016

Superior Set M:

SPY			
Rank of Models	v_M	v_R	Loss
HARQF error	-19.6898	-19.6898	-0.000016
p-value:	0		
KO			
Rank of Models	v_M	v_R	Loss
HARQ error	-7.1199	-7.1199	-0.00000987
p-value:	0		
MSFT			
Rank of Models	v_M	v_R	Loss
SHAR error	-0.2415	2.0816	-0.000062
HARQ error	-2.0947	-1.7473	-0.000064
HARQF error	1.4554	1.7473	-0.000059
p-value	0.196		
AXP			
Rank of Models	v_M	v_R	Loss
HARQF error	-8.7727	-8.7727	0.000114
p-value	0		

Notes. For the MCS procedure an α of 0.10 is employed and the number of bootstrapped samples to construct the test statistic is 1000.

Table 11: Model Confidence Sets for all four datasets