Evaluation of superior prediction of GARCH models for a volatility proxy

Andrew Kolenchenko(560495), Steinar Horst(575922), Simon Benhaiem(572172), Akram Aykour(560202), team X5

June 2022

Word Count: 4968

Abstract

Investors attempt to forecast the market volatility accurately as it plays an important role in financial investment decisions. We focus on the predictive ability of the general autoregressive conditional heteroskedastic (GARCH) models, relative to the benchmark models, the VIX and the heterogeneous autoregressive-realized volatility (HAR-RV). Our present study employs a data set consisting of close-to-close S&P 500 returns, realized variances, and the VIX from January 2000 to May 2022. We find contradicting evidence that the GARCH models do not outperform the benchmark models and that especially HAR-RV remains superior.

1 Introduction

Modelling present and future stock market volatility is a popular agenda in financial markets which has many applications in risk management, asset pricing and portfolio selection. Researchers have made substantial progress in modelling the time variation of volatility. Before 1982, conservative econometric models assumed a constant one-period ahead forecast variance. Since this assumption is outdated and too simplistic, Engle (1982) proposed the autoregressive conditional heteroskedastic (ARCH) processes. At a later point, Bollerslev (1986) introduces the generalized ARCH model (GARCH).

The current paper aims to investigate whether GARCH models outperform VIX or HAR-RV in forecasting volatility proxies. We measure the quality of the prediction performance using out-of-sample forecasting analysis using three different robust loss functions (Patton, 2011) and volatility proxies based on squared returns or high frequency returns. Our results are contradictory to Hansen & Lunde (2005) that concluded that the GARCH(1,1) model performs the best. In our research paper we provide the reader with empirical evidence that the HAR-RV, instead, performs the best.

The paper commences with Section 2, which introduces the reader to the necessary background information surrounding forecasting the market volatility. We then introduce and analyse the given data set used in the research paper, within Section 3. Section 4 entails the description of the estimation, forecasting and statistical comparison procedures. The corresponding results in Section 5 are then represented graphically and in tabular data, with their corresponding descriptions. Furthermore, we summarise our important findings and evaluate our paper to construct the conclusion in Section 6. Finally, we include an Appendix in Section 7 which contains the remaining results and derivations.

2 Theory

Accurately predicting the market volatility is a central topic in finance which would allow investors to make informed financial decisions. The market volatility is unobservable, which leads to the use of proxies that attempt to capture its effect. We use two of such proxies: the close-to-close squared returns, and the realised variance constructed from intra-day returns with a five-minute interval. The advantage of the second proxy is that it captures more data by considering intra-day volatility instead of only the inter-day volatility. If within a trading day the market is very volatile but settles down to the previous closing at the end of the day (i.e. close-to-close), the first proxy would indicate a steady market while the second proxy would show the occurred volatility. One caveat of the realised variance, is that it does not capture overnight volatility for which some heuristic adjustments can be made.

To predict and explain the volatility proxies there exist certain approaches. Brenner (1989) proposed the creation of a stock market volatility index (VIX). The VIX is a popular measure for the stock market's expectation of volatility based on S&P 500 index options. The VIX is made up of options with the price of each option reflecting market expectations (CBOE, 2021).

The generalized autoregressive conditional heteroskedasticity (GARCH) model proposed by Bollerslev (1986), aims to describe the stock market volatility using daily close-to-close returns. As high-frequency data of returns becomes more common and of better quality, a more intricate model could be more adept. The Realised GARCH model by Hansen et al. (2012) incorporates intra-day returns. Both of these GARCH models can be altered in such a way that it takes the direction of a sudden shock in the market into account. Black (1976) argues that negative market shocks affect the market differently than positive market shocks. The models that result from this alteration are so-called asymmetric GARCH models, which put different weights on the effect of a shock on the volatility depending on the direction. We perform analysis to choose between the asymmetric or symmetric variant. Furthermore, we consider further extensions of asymmetry in the asymmetric realised GARCH model (ARGARCH) which decomposes the model further. We will use the heterogeneous autoregressive realised volatility (HAR-RV) as a benchmark model, which models the proxy directly on past realised variances over multiple-day horizons. Furthermore, VIX will also be used as a benchmark model.

For forecasting we consider two procedures: fixed and expanding window. Fixed static forecasting uses a fixed set of observations to construct the model, to make out-of-sample forecasts. An expanding window forecast takes new information iteratively and uses it to improve the next forecast. The expanding window resembles a practical algorithm, which investors might use.

To evaluate the predictive ability, we apply a certain loss function to the forecasts, to tell us how biased our forecasts are. We perform statistical analysis for superior predictive ability (SPA) for the GARCH models against the benchmark models. For our second statistical analysis, we investigate the added benefit of the decomposed asymmetric realised GARCH model as opposed to the asymmetric realised GARCH model, to see whether introducing additional asymmetry improves the forecasts.

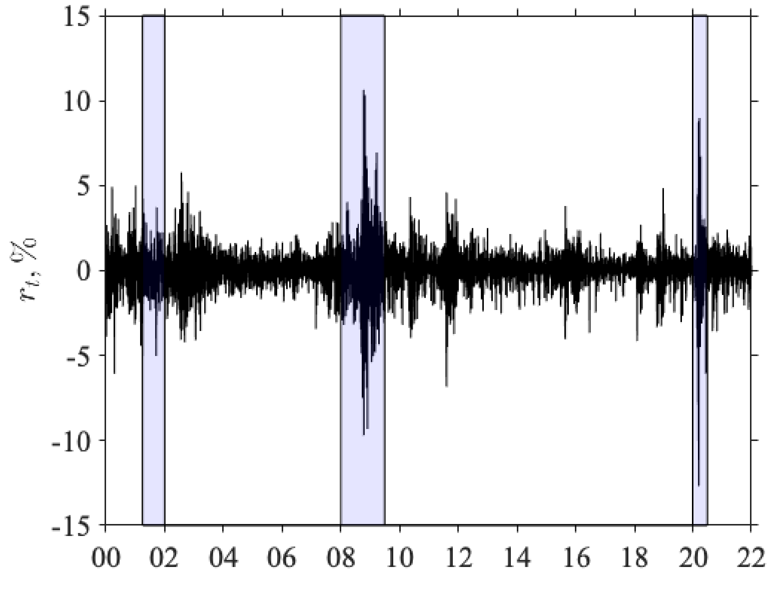
Based on Hansen & Lunde (2005), we hypothesize that the GARCH models have superior predictive ability compared to the benchmark models. Furthermore, we do not expect that the asymmetrically decomposed ARGARCH model will benefit predictive ability due to the principle of parsimony.

3 Data

We consider daily time series data from market returns for the last two decades. The time series contain close-to-close S&P 500 log returns from the 4th of January 2000 until the 20th of May 2022, resulting in 5610 observations. Alongside the returns, we have the realised variance data that incorporates five-minute intra-day returns collected from the Oxford Library. Furthermore, the daily VIX time series is considered which is provided by Chicago Board Options Exchange (CBOE).

Figure 1 represents the time series of the log returns, in which we do not observe a pronounced trend. We investigate the presence of a trend further by regressing the log return on a constant and deterministic trend coefficient. The corresponding Ordinary Least Squares (OLS) estimate for the trend coefficient is statistically insignificant considering a 10% significance level.

Figure 1: Returns with respect to the date (years) for close-to-close data. The shaded areas are National Bureau of Economic Research (NBER) quarterly recession periods



Furthermore, we do not observe a pronounced seasonal pattern within Figure 1. We can investigate further by regarding the following regression:

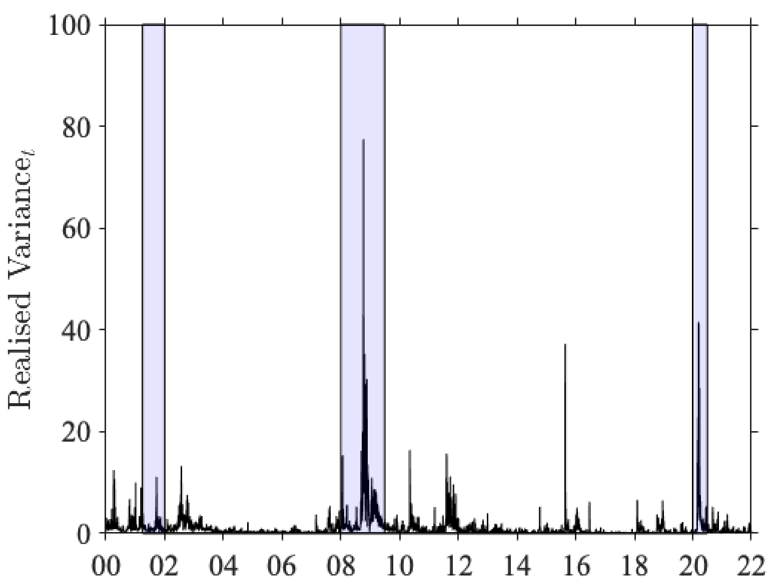
$$\Delta_1 r_t = r_t - r_{t-1} = \mu_1 D_{1,t} + ... + \mu_S D_{S,t} + \varepsilon_t \quad t = 2, 3, ..., T, \quad S = \{4, 12\}, \tag{3.1}$$

where $D_{s,t}$ takes the value 1 if the observation at time t is for quarter/month S and is equal to 0 otherwise. Note that if $S = 4$ (12), the regression is related to quarterly (monthly) seasons. A formal $F-$test of the H_0: $\mu_1 = ... = \mu_S$ gives p-values ≥ 0.99 for both seasons, indicating that there is no empirical evidence of seasonality in the original time series.

Figure 1 suggests that log returns tend to be more volatile during periods of economic recessions, while it stabilizes during expansions. The standard deviation is equal to 6.78 during NBER[1] recessions and 1.24 during expansions, suggesting that heteroskedasticity is present in the time series. Hence, we account for heteroskedasticity in the candidate models.

Secondly, we focus on the realised variance, which is a backward-looking volatility measure that is based on high-frequency returns. Figure 2 depicts the time series of the realised variance, in which we also do not observe a deterministic trend or seasonality. It is suggested by regressing realised variance on the trend and then on seasonal dummies like in Equation (3.1), where all coefficients obtain values that are close to zero. Moreover, there is a visible relation between the recession periods and the realised variance observations. There are clear observational peaks during the recession periods for the 2008 financial market collapse, for which the realised variance reaches a maximum. A peak in realised variance also occurs during the recession period for the COVID-19 pandemic in 2019-2020, which can be considered as an outlier.

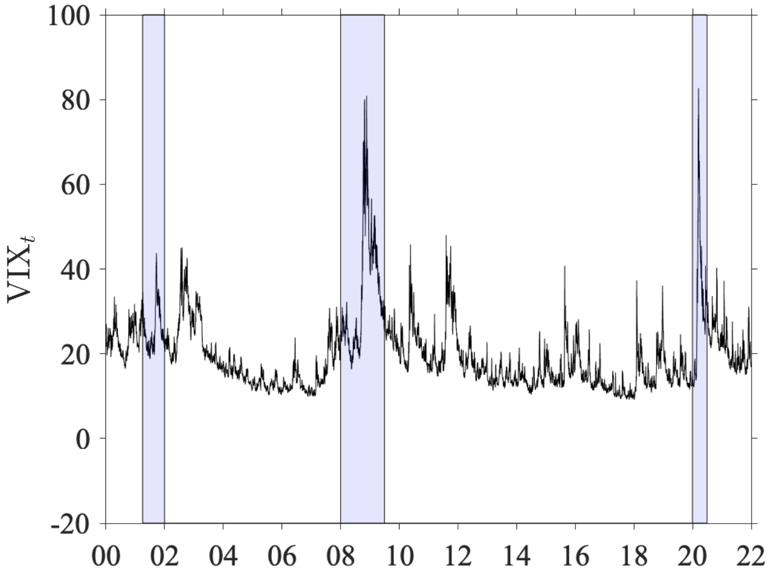
Figure 2: Realised Variance with respect to the date (years) for close-to-close daily data. The shaded areas are NBER quarterly recession periods



[1]Based on the turning points provided at http://www.nber.org/cycles/

In contrast with realised variance, the VIX is a forward-looking volatility measure which is based on the expectation of the future returns. Trend and seasonality are also not visible in Figure 3. Furthermore, the largest peak occurs during the recession of the COVID-19 pandemic, compared to the realised variance which reaches its peak during the 2008 financial market crisis.

Figure 3: VIX with respect to the date (years) for close-to-close daily data. The shaded areas are NBER quarterly recession periods



4 Methodology

Within Section 4 we discuss the procedure for estimating and forecasting the volatility proxy. Furthermore, we apply statistical inference to test whether the GARCH models outperform the benchmark models for predicting volatility proxies. We divide the elements of this procedure into subsections. In Subsection 4.1 we discuss the models that are used for predicting. Then, in 4.2 the target variables needed to evaluate our forecasts are defined. Furthermore, in 4.3 we explain how to estimate the models using the estimation sample. Moreover, the loss functions to evaluate the forecasts are specified in 4.4. Then, in 4.5 we describe both the forecasting procedures used. Lastly, we discuss the statistical procedure we apply to compare the performance measures of the forecasts in 4.6.

4.1 The candidate models

The models we use are divided in candidate models and benchmark models. First, we discuss the candidate models, which are based on the GARCH(1,1) and Realised GARCH models proposed by Bollerslev (1986) and Hansen et al. (2012), respectively. All the GARCH models used in this paper can in general be formulated as follows for all $t = 1, ..., 5610$:

$$r_t = \mu + \sigma_t \varepsilon_t, \tag{4.1}$$

$$\sigma_{t+1}^2 = \omega + \alpha (r_t - \mu)^2 + \beta \sigma^2 + \gamma x_t, \tag{4.2}$$

$$x_t = \xi + \varphi \sigma_t^2 + \tau(\varepsilon_t) + u_t, \tag{4.3}$$

$$\alpha = \alpha_1 \mathbb{1}_{\{\varepsilon_t < 0\}} + \alpha_2 \mathbb{1}_{\{\varepsilon_t \geq 0\}}, \tag{4.4}$$

$$\tau(\varepsilon_t) = \tau_1 \varepsilon_t + \tau_2 (\varepsilon_t^2 - 1), \tag{4.5}$$

where r_t represents returns, x_t stands for realised variance and σ_t^2 is the variance of r_t. For obtaining the regular symmetric GARCH model, we need to set $\gamma = 0$ and $\alpha_1 = \alpha_2$. To obtain the asymmetric variant we alleviate the restriction for α_1 and α_2 in (4.4). When we want to implement realised variance we allow γ to be non-zero, while keeping $\alpha_1 = \alpha_2$, obtaining the Realised GARCH model as described by Hansen et al. (2012). To get the symmetric GARCH version we set $\tau_1 = 0$, else we get the asymmetric Realised GARCH model (ARGARCH).

So, we have four nested models for GARCH in total. Out of these, we choose one standard GARCH and one Realised GARCH model, to avoid forecasts with too similar models. We choose by looking at the Akaike information criterion (AIC) and the Schwarz information criterion (SIC) for each model and compare.

As a preamble to our estimation procedures, we estimate (using the estimation procedure as described in Subsection 4.3) all four possible models using the full sample and compute the information criteria. In Tables 6 and 7 in subsection 7.3 of the Appendix, we see that both for the GARCH and the Realised GARCH the information criteria indicate that the asymmetric variant is lower, hence we prefer the asymmetric variant. We will therefore only use the AGARCH and ARGARCH as our candidate models. Using the asymmetric models means that we do take the direction of a shock into account, which goes in line with the leverage effect Black (1976) proposed.

As mentioned in Section 2 we also consider an asymmetric Realised model additionally decomposed for asymmetry. We decompose the original ARGARCH model by allowing not only α to be different for a positive or negative shock but also γ. Thereby, we obtain what we would like to call a decomposed asymmetric model. The derivations for obtaining the properties of this model are described in the Appendix, Subsection 7.3.

To rate the AGARCH & ARGARCH forecasting performance, we also include two benchmark models to give us an indication of the relative quality of a GARCH-type model. We introduce the VIX and the HAR-RV models, for which a more detailed explanation is provided in Subsection 4.5.

4.2 Target variables

As volatility cannot be directly observed, we consider two proxies of the market volatility. The first proxy uses close-to-close log returns, while the second is constructed using realised variance of intra-day returns.

The first proxy we will denote by TV1 and is defined as follows:

$$\text{TV1}_{t,d} = \sum_{i=1}^{d} r_{t+i}^2 \quad \text{for } t = 1, ..., 5589 \text{ and } d \in \{1, 5, 21\}, \tag{4.6}$$

where we sum the squares of the returns up until a specific horizon of trading days.

The second proxy is based on the realised variance that is denoted by TV2. As a mean average of variance during a trading day is approximately 70% and intra-day realised variance data does not take into account overnight trading activity, TV2 is upscaled by a factor $1/0.7 \approx 1.4$ (Hansen & Lunde, 2005) and is defined as follows:

$$\text{TV2}_{t,d} = 1.4 \times \sum_{i=1}^{d} \text{rv5\_ss}_{t+i} \quad \text{for } t = 1, ..., 5589 \text{ and } d \in \{1, 5, 21\}. \tag{4.7}$$

4.3 Estimation procedure

We use Maximum Likelihood Estimation (MLE) for both the AGARCH and ARGARCH as is described for the GARCH model by Bollerslev (1986). Retaining the assumptions of Hansen et al. (2012), most important of which are $\varepsilon_t \overset{i.i.d}{\sim} N(0,1)$ and $u_t \overset{i.i.d}{\sim} N(0, \sigma_u^2)$, the ML estimator for the full 10-dimensional parameter vector $\boldsymbol{\theta} = (\mu,\, \omega,\, \alpha,\, \beta,\, \gamma,\, \xi,\, \varphi,\, \tau_1,\, \tau_2,\, \sigma_u)$ is optimised within Equation (4.8) that is based on the probability distribution function of error terms ε_t.

$$\hat{\boldsymbol{\theta}}_{\text{ML}} = \text{argmax}_{\boldsymbol{\theta}} \underbrace{\sum_{t=1}^{n} \frac{1}{2}\Bigg[-\log(2\pi) - \log(\sigma_t^2) - \frac{(r_t-\mu)^2}{\sigma_t^2}\Bigg]}_{l(r)} + \underbrace{\sum_{t=1}^{n} \frac{1}{2}\Bigg[-\log(2\pi) - \log(\sigma_u^2) - \frac{u_t^2}{\sigma_u^2}\Bigg]}_{l(x|r)} \tag{4.8}$$

As we have a high-dimensional non-linear optimisation problem, the specification of the starting values play a sensitive role in the convergence to an optimal solution. Hence, a pragmatic approach is used. For μ a sample mean of returns is chosen as an initial value (μ_0), while the starting point for ω, (ω_0) is the variance of returns divided by 20. Furthermore, for the remaining parameters, we choose a vector of starting points defined as follows: $(\alpha_0,\, \beta_0,\, \gamma_0) = (0.10,\, 0.85,\, 1)$. Firstly, we optimise with respect to (μ, ω, α, β, γ), which corresponds to the $l(r)$ part of Equation (4.8). We use the estimates from the first optimisation problem to compute the $\hat{\sigma_t}^2$ and $\hat{\varepsilon_t}^2 \quad \forall t$. Then we regress x_t on the dependent variables of Equation (4.3) using OLS to obtain the remaining starting values $\hat{\boldsymbol{\theta}}^{\text{OLS}} = (\hat{\xi}^{\text{OLS}}, \hat{\varphi}^{\text{OLS}}, \hat{\tau_1}^{\text{OLS}}, \hat{\tau_2}^{\text{OLS}}, \hat{\sigma_u}^{\text{OLS}})$, where $(\hat{\sigma_u}^{\text{OLS}})^2$ is a mean squared error of a regression. The entire 10-dimensional optimisation in Equation (4.8) with the starting values $\hat{\boldsymbol{\theta}}_0 = (\mu_0,\, \omega_0,\, 0.01,\, 0.5,\, \gamma_0,\, \hat{\boldsymbol{\theta}}^{\text{OLS}})$ are used to obtain adequate estimates. Note, by decreasing the values of α_0 and β_0 we are preventing an initial point problem and hence satisfying the function barrier $\alpha + \beta + \gamma\phi < 1$.

4.4 Loss functions

To be able to quantify the forecasting performance of both the GARCH type models and the benchmark models, we use three distinct loss functions. The loss functions aim to penalize biased forecasts, by comparing the predicted variance with the target variances. The three loss functions all come from Equation (4.9), which is a family of functions proposed by Patton (2011), shown to be robust and homogeneous loss functions for forecasts of a volatility proxy. Other measures like MAE are viable for other applications but as Patton (2011) shows, these are not robust when using volatility proxies. The following piece-wise function displays said family of functions:

$$L(\hat{\sigma}^2, h; b) = \begin{cases} \frac{(\hat{\sigma}^{2b+4} - h^{b+2})}{(b+1)(b+2)} - \frac{h^{b+1}(\hat{\sigma}^2 - h)}{b+1} & \text{for} \quad b \notin \{-1, -2\} \\ h - \hat{\sigma}^2 + \hat{\sigma}^2 \log(\frac{\hat{\sigma}^2}{h}) & \text{for} \quad b = -1 \\ \frac{\hat{\sigma}^2}{h} - \log(\frac{\hat{\sigma}^2}{h}) - 1 & \text{for} \quad b = -2 \end{cases}, \tag{4.9}$$

where h is our predicted variance and $\hat{\sigma}^2$ is our proxy for volatility. We specifically use $b = -2$, $b = 0$, and $b = 1$ which is described in the following equations:

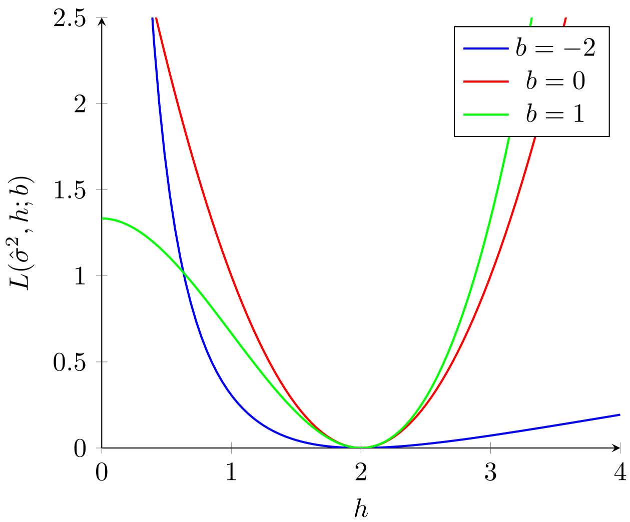
$$L_{-2} := L(\hat{\sigma}^2, h; -2) = \frac{\hat{\sigma}^2}{h} - \log(\frac{\hat{\sigma}^2}{h}) - 1, \tag{4.10}$$

$$L_0 := L(\hat{\sigma}^2, h; 0) = (\hat{\sigma}^2 - h)^2, \tag{4.11}$$

$$L_1 := L(\hat{\sigma}^2, h; 1) = \frac{\hat{\sigma}^6 - h^3}{6} - \frac{h^2(\hat{\sigma}^2 - h)}{2}. \tag{4.12}$$

Note that Equation (4.11) should be divided by two if we strictly follow the family of functions in Equation (4.9). We choose to multiply Equation (4.11) by two, to get the popular Mean-Squared Error (MSE) loss function. Equation (4.10) corresponds with the QLIKE function.

Figure 4: Robust loss functions with true $\hat{\sigma}^2 = 2$



In Figure 4 we see the mathematical properties of the three proposed loss functions. We see that the L_0 is a symmetric function (around the true $\hat{\sigma}^2$ 'proxy'), which means that it penalizes undervalued and overvalued predictions to the same degree. The L_{-2} and the L_1 loss functions are asymmetric (around the true $\hat{\sigma}^2$ 'proxy'), which means that they do not penalize evenly. Specifically, L_{-2} penalizes the overvalued predictions less as opposed to L_1, which is biased towards undervalued predictions. In general, a lower value for L_{-2}, L_0, L_1 means a more favoured prediction, given the inherent bias of the loss function. Especially, when the loss function obtains a value of zero for a certain prediction, it means that the prediction is equal to the actual volatility proxy.

When we look at the symmetrical properties of the three loss functions, we can consider the economic interpretation of using each function. Firstly, as L_0 is symmetric, we can associate it with risk-neutrality as the function does not place any bias towards undervalued or overvalued prediction. Secondly, as the L_{-2} function penalizes underpredicting volatility more heavily, the comparisons using this function are more biased towards 'safer' forecasts. The model picked by L_{-2} will be preferred by risk-averse agents. Lastly, the L_1 loss function penalises in the reverse way as the L_{-2} function, by preferring forecasts that underpredict the volatility and thus minding an unexpectedly high level of volatility less. The model picked by L_1 will be preferred by risk-seeking agents.

4.5 Forecasting procedure

After having obtained the optimal parameter estimates we proceed with forecasting the target variables. We do this in two ways. Firstly, we use a fixed window to estimate the parameters, predicting the target variable one-step ahead. Secondly, we make use of an expanding window, estimating the model repeatedly, by adding one data point to the estimation sample each time after forecasting.

For the fixed window forecast we use 50% of the data, 2805 observations, to estimate the parameters in the model(s) described in Equations (4.1) to (4.5). Then, we predict the remainder of the time series.

For the expanding window forecast, we start with the same 50% of the data. Then we make a d-day ahead forecast, $d \in \{1, 5, 21\}$. For the next forecast, we expand the estimation sample by five observations. The choice of the weekly step size is two-fold. On one hand, we want to minimize the step size on account of a smaller step size making the estimations close to the days we predict for. On the other hand, having a big step size decreases computation time.

Note, we leave out the last 25 observations to retain the same amount of observations predicted for each horizon. The target variable for the 21-day horizon cannot be computed for the last 21 observations, so we leave out the last 21 observations. Furthermore, as we have a step size of five and we want to make each forecast using the same step size, we leave out four more observations.

The point forecasts themselves are based on minimising the loss functions mentioned within Subsection 4.4. The minimisation of the loss functions results in the optimal point forecast being the conditional expectation of the estimated variance (Patton, 2011). The point forecast of $\hat{\sigma}^2_{T+h}$ is denoted as $\hat{\sigma}^2_{T+h|T}$ where d denotes the forecast horizon for the forecast origin time T.

The point forecast for the estimated proxy volatility can be described as follows:

$$\hat{\sigma}^2_{T+d|T} = \mathbb{E}(\sigma^2_{T+d}|I_T), \tag{4.13}$$

where I_T is the information set at time T, that is $I_T = \{\sigma^2_{T+1}, \cdots, \sigma^2_{T+d-1}\}$. For evaluation of the forecasts by comparing them to the described target variables in Section 4.2, the point forecasts are used for constructing the predicted variances (PVs). The formulation for the PV of each model is described below.

The general formulation for the predicted variance (PV) for ARGARCH is formulated as

$$\mathrm{PV}_{t,d} := \mathbb{E}_t\left[\sum_{i=1}^{d} r_{t+i}^2\right] = d(\mu^2+\sigma^2) + \frac{1-(\alpha+\beta+\gamma\varphi)^d}{1-\alpha-\beta-\gamma\varphi}(\sigma_{t+1}^2-\sigma^2) \qquad d\in\{1,5,21\}, \tag{4.14}$$

where

$$\sigma^2 := \frac{\omega+\gamma\xi}{1-\alpha-\beta-\gamma\varphi} \qquad \text{assuming} \qquad \alpha+\beta+\gamma\phi<1 \tag{4.15}$$

Note that we can obtain the $\mathrm{PV}_{t,d}$ for the AGARCH model by setting $\gamma = 0$ and allowing $\alpha = \frac{\alpha_1+\alpha_2}{2}$.

The VIX aims to predict the annualised standard deviation, which means we have to square and scale the VIX series to be able to compare it with the GARCH type models. The predicted variance at time t over a d-day horizon is computed as follows:

$$\mathrm{PV}_{t,d} = d/250 \times \mathrm{VIX}_t^2, \tag{4.16}$$

where we scale the VIX_t^2 by 1/250 to account for the annualization of the effect of the month-ahead expectation. Furthermore, we scale by d to obtain the effect for the wished d-day ahead forecast.

The HAR-RV, as proposed by Corsi (2009), works by first regressing $\mathrm{TV}_{t,d}$ onto a constant with three realised variance components, and then computes the predicted variances using OLS estimates. We use the formulas and definitions as described in Corsi (2009).

4.6 Statistical tests

To give a statistical judgement to the predictive abilities of our model forecasts, we closely look at two tests and their assumptions, while keeping the multiple inferences problem in mind.

As for the statistical procedure we closely follow Li & Patton (2018), who provide us with a general framework for statistical inference methods of predictive ability in applications such as volatility forecasting. Li & Patton (2018) specifically focus on the fact that we use proxies for volatility (as volatility is unobservable), but that we still want to test predictive superiority based on the true volatility. Such a 'true' hypothesis is infeasible, and therefore requires us to make a 'proxy' hypothesis (Li & Patton, 2018) (i.e. a hypothesis containing the proxy instead of the true volatility value). The question is how good such proxy hypotheses are in representing the true hypothesis, specifically how quickly the proxy hypothesis asymptotically converges to the true hypothesis. The convergence-in-hypothesis is closely related to the convergence rate of the proxy toward the true value[2]. Li & Patton (2018) show that many existing equal (superior) predictive ability (EPA/SPA) testing procedures such as the ones described in Diebold & Mariano (1995), West (1996), White (2000) and Hansen (2005), attain the asymptotic negligibility result[3].

[2]Note that Li & Patton (2018) specifically talk about high-frequency proxies that converge to the target value, as the sampling interval (of the high-frequency data) goes to zero

[3]Li & Patton (2018) describes this as forecast evaluation methods using proxies that have the same asymptotic size and power properties under the proxy hypotheses as under the true hypotheses

We will use two types of SPA testing procedures. First, we introduce a joint SPA test as described in Hansen (2005), and secondly, we introduce a Multi-Horizon SPA test as described in Quaedvlieg (2021) which is based on the Diebold-Mariano test in Diebold & Mariano (1995). Note that the joint SPA test in Hansen (2005) is shown to attain the asymptotic negligibility result in Li & Patton (2018), while the Multi-Horizon SPA test (as it was published later) was not shown to attain the desired result. We do decide to use the Multi-Horizon SPA test as it is a modified Diebold-Mariano test that was shown to attain the asymptotic negligibility.

Both the tests contain the same setup. Let $h_{i,t}^d$ be the point forecast at multiple horizons $d = 1, ..., D$. That is, for model $i = 1, .., M$, we have the forecasts $h_{i,t} = [h_{i,t}^1, h_{i,t}^2, ..., h_{i,t}^D]'$. We then use the loss functions $L_{i,t}(\hat{\sigma}_t^2, h_{i,t}; b)$ from Section 4.4, which gives us a D-dimensional vector with elements $L_{i,t}^d(\hat{\sigma}_t^2, h_{i,t}^d; b)$. For any specified loss function and two forecasts (from two different models), we then define the loss differential between two models as:

$$d_{ij,t} := L_{i,t}(\hat{\sigma}_t^2, h_{i,t}; b) - L_{j,t}(\hat{\sigma}_t^2, h_{j,t}; b). \tag{4.17}$$

As we have three horizons, three loss functions, five models, two target variables and two forecasting procedures, we have many opportunities to fall into the multiple inference problem (Tukey, 1953). To combat the statistical problem we can control the probability of making at least one type I error in the family of hypotheses (FWER). To control the FWER we use the Bonferroni correction (Bonferroni, 1936), which 'corrects' the significance level by dividing the pre-specified significance level (e.g. 5%) by the number of inferences being done on our data set.

4.6.1 Joint SPA

We choose to use the 'joint' SPA test of Hansen (2005) as we want to test whether any model (out of a basket of models) significantly outperforms a specified benchmark model in terms of forecasting ability. We will use the joint SPA test to test whether the AGARCH or ARGARCH model can outperform either the VIX or HAR-RV benchmark model for a given loss function (L_{-2}, L_0, L_1), a given horizon ($d \in \{1, 5, 21\}$), a given target variable (TV1 or TV2), and a given forecasting procedure (fixed or expanding window). This gives us 72 different tests[4], which is for many reasons too many as we are guaranteed to fall into the bias of data snooping (White, 2000). The joint SPA test is based on the Reality Check test as described in White (2000), which means that it also assesses the dangers of data snooping. We decide to only look at 1 specific loss function, and 1 target variable, which means we still have 12 tests to perform. We choose TV2 from Equation (4.7) as it incorporates intra-day returns as opposed to TV1 which uses close-to-close returns. We also choose the QLIKE loss function, since we are interested in the risk-averse perspective (as described in 4.4), in light of the recent increase in market volatility such as the COVID crisis (described in Section 3). For the testing procedure we define $d_{i,t} := d_{0i,t}$ from Equation (4.17) where "0" denotes the specified benchmark. $d_{i,t}$ denotes the performance

[4] 2 benchmark model × 3 loss functions × 3 horizons × 2 target variables × 2 forecasting procedures = 72

of model i relative to the benchmark model at time t. By stacking over the different models, we obtain $d_t = [d_{1,t}, ..., d_{M,t}]'$. By taking the mean $\mu^{SPA} := \mathbb{E}[d_t]$, we can formulate the null hypothesis as:

$$H_0 : \mu^{SPA} \leq 0, \tag{4.18}$$

with alternative $\mu^{SPA} > 0$. The alternative refers to that at least one of the models is better than the benchmark model in terms of expected loss. See Hansen (2005) for more information on the exact implementation of the test.

4.6.2 Multi-Horizon SPA

The Multi-Horizon SPA test described in Quaedvlieg (2021) enables us to jointly compare forecasts of different models across different forecasting horizons, rather than individually comparing forecasts at different horizons. Quaedvlieg (2021) introduces two such tests: the uniform SPA and the average SPA test. The first allows us to test for superior performance at every horizon, while the second test allows for inferior performance at some of the horizons. The uniform SPA test is stricter because, to conclude uniform SPA, we need for all horizons a superior forecasting performance. Quaedvlieg (2021) shows us, using simulation studies, that the described tests contain appropriate size and high power.

For the testing procedure we use $d_{ij,t}$ as defined in (4.17). The hypothesis is based on the mean of the expected loss differential, $\mu_{ij} := \lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} \mu_{ij,t}$ where $\mu_{ij,t} = \mathbb{E}[d_{ij,t}]$

Note that we can construct the Diebold-Mariano test by investigating the null hypothesis:

$$H_{DM} : \mu_{ij}^{d} = 0, \tag{4.19}$$

at a single horizon d, and by using a standard t-test.

For the test for Multi-Horizon average SPA, we define the weighted average loss difference as follows:

$$\mu_{ij}^{(\text{Avg})} = \sum_{d=1}^{D} w_d \mu_{ij}^{d}, \tag{4.20}$$

with the positive weights w_d summing up to one. We use equal weighting for our horizons, as we do not 'prefer' any particular horizon. The corresponding null hypothesis is

$$H_0 : \mu_{ij}^{(\text{Avg})} \leq 0, \tag{4.21}$$

with alternative $\mu_{ij}^{(\text{Avg})} > 0$. The alternative refers to model j outperforming model i. The null is then tested with a simple t-test. The test is made such that the compared models can be nested or non-nested (unlike Diebold & Mariano (1995)), but that the forecasting procedure can only be fixed or rolling window. Thus, we will only test average SPA for the fixed window forecasts. See Quaedvlieg (2021) for more information on the exact implementation of the test. For our implementation of the test, we compare the average multi-horizon SPA of the decomposed ARGARCH against the original ARGARCH using TV2 and all three loss functions. As the test jointly compares over horizons, we perform three such tests.

We perform 15 SPA tests and three t-tests in Section 3 resulting in 18 inferences. We use Bonferroni correction to adjust our prior significance level of 0.05 to $\frac{0.05}{18}$ ≈0.0028.

5 Results

The results of the forecasts are divided into two parts. In Subsection 5.1 we consider the main results concerning the research question, while in the second part we discuss the merits of the asymmetrically decomposed ARGARCH model.

5.1 GARCH versus benchmarks

First, we consider our main test for comparing forecast performance. The result of the joint SPA tests are displayed in Table 1. For expanding window forecasting, which gives GARCH the best chance in forecasting, we find that for the comparison between GARCH models and VIX we reject twice at our Bonferroni corrected significance level of 0.0028. This means that for two of the three horizons the GARCH outperforms the VIX in terms of forecasting performance. For the comparison against HAR-RV, we can only find one instance where the GARCH models perform better. This tells us that the GARCH models might have a chance against the VIX but are not adept enough to beat HAR-RV.

Table 1: p-values for the joint SPA test for comparing GARCH models to the benchmark models over the three horizons for TV2 and QLIKE

	Fixed static			Expanding Window		
	d = 1	d = 5	d = 21	d = 1	d = 5	d = 21
to VIX	<0.001*	0.004	0.004	<0.001*	0.004	0.002*
to HAR-RV	>0.999	>0.999	0.020	>0.999	>0.999	<0.001*

$*$:= significant at Bonferroni corrected significance level of ca. 0.0028

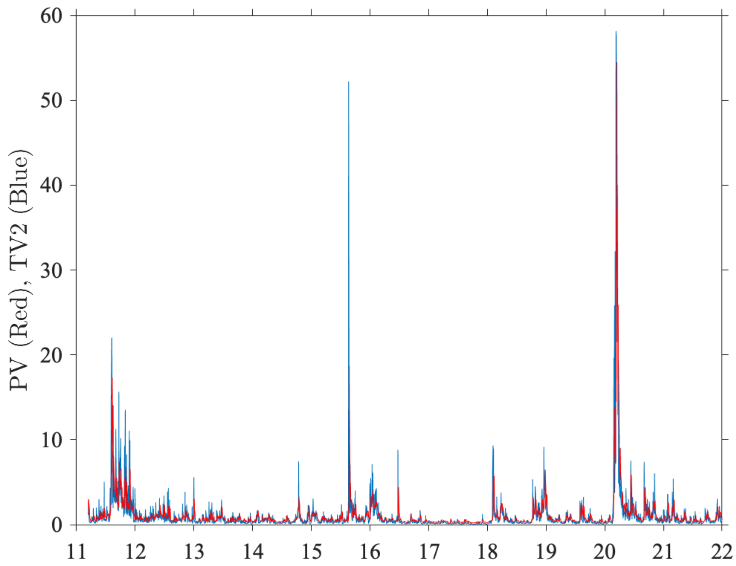
Besides considering the statistical comparison we have an overview of all performance measures, i.e. the mean of the loss functions, for the volatility proxy forecasting. In Table 2 we provide the MSE, QLIKE and $\bar{L}_1$ values for the expanding window forecasts. The table for the fixed window can be found in the Appendix in 7.5, as the structure is similar and we prefer the economic perspective for the expanding window forecasts.

Table 2: MSE, QLIKE and $\bar{L}_1$ values for the expanding window forecasts (with minimum values per TV per loss function bolded)

	Panel A: One day ahead (d = 1)							
	TV1				TV2			
	AGARCH	ARGARCH	VIX	HAR-RV	AGARCH	ARGARCH	VIX	HAR-RV
MSE	19.470	18.068	20.077	**17.364**	6.904	4.459	4.779	**4.275**
QLIKE	1.625	1.557	1.670	**1.505**	0.363	0.309	0.420	**0.278**
$\bar{L}_1$	408.533	379.973	393.058	**363.940**	80.637	41.181	**32.932**	37.943
	Panel B: One week ahead (d = 5)							
	TV1				TV2			
	AGARCH	ARGARCH	VIX	HAR-RV	AGARCH	ARGARCH	VIX	HAR-RV
MSE	279.779	232.029	244.017	**224.238**	160.365	83.511	81.899	**75.304**
QLIKE	0.541	0.489	0.529	**0.447**	0.325	0.289	0.338	**0.256**
$\bar{L}_1$	19575.196	15360.715	15177.716	**14600.600**	9374.497	3121.112	**1524.903**	2403.520
	Panel C: One month ahead (d = 21)							
	TV1				TV2			
	AGARCH	ARGARCH	VIX	HAR-RV	AGARCH	ARGARCH	VIX	HAR-RV
MSE	4515.598	3204.128	3203.316	**3135.480**	3321.050	1606.820	1372.897	**792.404**
QLIKE	0.489	0.488	**0.438**	3.904	**0.375**	0.394	0.473	2.466
$\bar{L}_1$	834711.089	407691.830	**388387.714**	391018.667	688884.949	159336.743	**18057.959**	50667.251

Generally, we can see from Table 2 that as the time horizon increases, the predictive ability for all models decreases substantially, which is reasonable to expect. Moreover, we see that in general $\bar{L}_1$ is relatively high indicating that our forecasts are underpredicting more so than overpredicting. This is illustrated in Figure 5, where we display an example of one of our forecasts and we see that the red line indicating the forecast is generally lower than the actual proxy.

Figure 5: ARGARCH for TV2, d = 21,



Another interesting observation is that all the forecasts seem to improve when we consider TV2 instead of TV1, strengthening our reason of choosing for TV2 as the more reliable volatility proxy. This improvement is interesting since AGARCH does not use realised volatility in its estimation and forecasts but still performs much better when using realised variance as a comparative proxy.

Similar to the results from the joint SPA test, we observe that HAR-RV has the lowest values overall, making it the best predictor. However, we do see that the GARCH models come close to HAR-RV in the case of QLIKE, such as in Table 2, Panel B for target variable 2, where ARGARCH attains a value of 0.289 compared to 0.256 for HAR-RV. Even better, the ARGARCH model outperforms the HAR-RV in Panel C, TV2 where the ARGARCH attains a QLIKE of 0.375 versus HAR-RV which attains a QLIKE of 2.466. This result seems to indicate that HAR-RV gets worse for bigger horizons, which could be due to it backward-looking model structure.

In smaller horizons, GARCH models generally attain lower measures in relation with VIX. When we increase the horizon, however, the VIX seems to outperform the GARCH models for most loss functions. This could mean VIX is more stable, which makes sense when we consider that it is a forward-looking predictor and is therefore, less affected by previous data like the GARCH models. Note that this is not in line with the joint SPA test which generally supports the hypothesis that the GARCH models outperform VIX, even for high horizons. This result is probably due to us choosing QLIKE for the statistical analysis, which is confirmed by Table 2, where we see that GARCH models outperforms VIX mostly for QLIKE.

5.2 Asymmetric adjustment

We test the asymmetrically decomposed ARGARCH model using the Average Multi-Horizon SPA test, as described in Subsection 4.6.2, for all performance measures using TV2.

Table 3: Average Multi-horizon SPA test for the ARGARCH model versus the asymmetrically decomposed ARGARCH

	Test statistic	p-value
MSE	-1.153	> 0.999
QLIKE	0.456	0.100
$\bar{L}_1$	-1.139	> 0.999

Even without the Bonferroni correction we do not reject the hypothesis of SPA along averaged horizons for the decomposed model, as illustrated in Table 3.

Table 4: MSE, QLIKE and $\bar{L}_1$ values for ARGARCH and its asymmetrically decomposed counterpart (with minimum values between panels bolded)

ARGARCH						
	One day ahead d = 1		One week ahead d = 5		One month ahead d = 21	
	TV1	TV2	TV1	TV2	TV1	TV2
MSE	18.124	**4.437**	**232.925**	**83.342**	**3183.002**	**1602.722**
QLIKE	**1.557**	**0.312**	**0.491**	0.295	0.494	0.411
$\bar{L}_1$	**379.432**	**39.985**	**15281.266**	**2986.192**	**399766.817**	**150558.393**
ARGARCH additionally decomposed for asymmetry						
	One day ahead d = 1		One week ahead d = 5		One month ahead d = 21	
	TV1	TV2	TV1	TV2	TV1	TV2
MSE	**17.687**	5.467	241.035	120.811	3718.300	2438.758
QLIKE	1.560	0.333	0.495	**0.294**	**0.466**	**0.372**
$\bar{L}_1$	385.897	68.159	17763.125	7335.254	624710.760	446525.585

This is further shown in Table 4, as the performance measures are not lower in general for the decomposed model. These findings suggest that it is not worth the effort to further decompose the ARGARCH model, since prediction accuracy does not increase significantly for the considered target variable.

6 Conclusions

The question we pose is whether GARCH models outperform the VIX and HAR-RV in predicting volatility proxies. Our results contradict the statement of Hansen (2005), that nothing beats the GARCH(1,1) model. We empirically show that the considered GARCH models, generally, do not outperform HAR-RV. However, the GARCH models outperform the VIX in some instances. Furthermore, we find that implementing the decomposed ARGARCH model is not worth the effort, concluding from the average Multi-Horizon SPA test.

For future applications, we suggest instrumental variables for the returns that do not directly influence stock market volatility but that are strongly correlated with the returns rates. As returns depend on a plethora of macroeconomic and financial factors, the endogeneity problem could potentially arise which could negatively affect the forecasts' precision. Explaining the returns by the instrumental variables, that satisfy 2-Stage-Least-Squares requirements and have high-frequency data, could improve the existing GARCH model. Therefore, further investigation is recommended. Moreover, the Bonferroni correction is the most conservative way to control for the multiple-inference problem. However, it ignores the distribution of the p-values across inferences, so we can consider alternative corrections such as Holm (1979) step-down procedure, which does account for the distribution of the p-values.

References

Black, F. (1976). Studies of stock market volatility changes. *1976 Proceedings of the American statistical association bisiness and economic statistics section*.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of econometrics*, *31*(3), 307–327.

Bonferroni, C. (1936). Teoria statistica delle classi e calcolo delle probabilita. *Pubblicazioni del R Istituto Superiore di Scienze Economiche e Commericiali di Firenze*, *8*, 3–62.

Brenner, M. (1989). New financial instruments for hedging changes in volatility, 45: 4 fin. analysts j. 61–65 (july/aug. 1989); menachem brenner & dan galai. *Hedging Volatility in Foreign Currencies*, *1*(1), 53–58.

CBOE. (2021). *White paper - chicago board options exchange.* Author. Retrieved from `https://cdn.cboe.com/resources/vix/vixwhite.pdf`

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

Diebold, F. X., & Mariano, R. S. (1995). Comparing predictive accuracy. *Journal of Business & Economic Statistics*, *13*(3).

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. *Econometrica: Journal of the econometric society*, 987–1007.

Hansen, P. R. (2005). A test for superior predictive ability. *Journal of Business & Economic Statistics*, *23*(4), 365–380.

Hansen, P. R., Huang, Z., & Shek, H. H. (2012). Realized garch: a joint model for returns and realized measures of volatility. *Journal of Applied Econometrics*, *27*(6), 877–906.

Hansen, P. R., & Lunde, A. (2005). A forecast comparison of volatility models: does anything beat a garch (1, 1)? *Journal of applied econometrics*, *20*(7), 873–889.

Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian journal of statistics*, 65–70.

Li, J., & Patton, A. J. (2018). Asymptotic inference about predictive accuracy using high frequency data. *Journal of Econometrics*, *203*(2), 223–240.

Patton, A. J. (2011). Volatility forecast comparison using imperfect volatility proxies. *Journal of Econometrics*, *160*(1), 246–256.

Quaedvlieg, R. (2021). Multi-horizon forecast comparison. *Journal of Business & Economic Statistics*, *39*(1), 40–53.

Tukey, J. W. (1953). The problem of multiple comparisons. *Multiple comparisons*.

West, K. D. (1996). Asymptotic inference about predictive ability. *Econometrica: Journal of the Econometric Society*, 1067–1084.

White, H. (2000). A reality check for data snooping. *Econometrica*, *68*(5), 1097–1126.

7 Appendix

7.1 Implemented feedback

We implemented the feedback provided by our peers. Firstly, we made sure to be more descriptive, regarding the meaning of the variables described in equations, as our peers suggested. Furthermore, some parts of the paper, especially the results and conclusions were in past tense, and as we believe that the results and conclusion should be presented as current information open to scrutiny, we changed the verbs in question to present tense, as the feedback suggested. Next, we added commas to the longer sentences, as suggested, in order to clear up the structure of these sentences. Lastly, we corrected all grammar mistakes found by our peers and by ourselves.

7.2 The estimation of GARCH parameters

The corresponding values in Table 5 is used for calculating the respective AIC and SIC values for the symmetric and asymmetric variants of the GARCH and realised GARCH models.

Table 5: Output table for the GARCH model and its extensions containing the parameter estimates and log-likelihoods

	S GARCH	AS GARCH	S R GARCH	AS R GARCH
$\hat{\mu}$	0.059	0.015	0.051	0.0381
$\hat{\omega}$	0.022	0.021	0.055	0.055
$\hat{\alpha}$	0.129	-	0.099	0.097
$\hat{\alpha_1}$	-	0.181	-	-
$\hat{\alpha_2}$	-	0.003	-	-
$\hat{\beta}$	0.859	0.89	0.522	0.520
$\hat{\gamma}$	-	-	0.466	0.469
$\hat{\xi}$	-	-	0.094	0.090
$\hat{\varphi}$	-	-	0.675	0.677
$\hat{\tau_1}$	-	-	-	-0.148
$\hat{\tau_2}$	-	-	0.184	0.161
$\hat{\sigma_u}$	-	-	1.617	1.612
log-likelihood	-7712.200	-7611.500	-18178.300	-18157.700
partial log-likelihood	-7712.200	-7611.500	-7536.800	-7536.800

7.3 Choosing between the GARCH models

When comparing the output of the parameter estimates for the GARCH models in Table 5, we see that the log-likelihood for the symmetric filter is higher than for the asymmetric filter. To be more careful we also consider the number of parameters estimated in each model: 4 in symmetric GARCH and 5 in the asymmetric variant. Combining these two components we can compute the corresponding Akaike Information Criterion (AIC) and Schwarz Information Criterion (SIC) based on Equations 7.1 and 7.2, respectively.

$$\text{AIC} = -2\log(\hat{L}_{\text{MLE}}) + 2k \tag{7.1}$$

$$\text{SIC} = -2\log(\hat{L}_{\text{MLE}}) + k\log(n) \tag{7.2}$$

The $\hat{L}_{\text{MLE}}$ denotes the maximum log likelihood value of the GARCH models. The number of parameters and observations within Equations 7.1 and 7.2 are denoted by k and n, respectively. The data needed to compute the respective AIC and SIC values can be found in Section 7, Table 5. The corresponding AIC and SIC values can be found in Table 6.

Table 6: The AIC and SIC values for the GARCH variants

	Symmetric GARCH	Asymmetric GARCH
AIC	15432.4	15233.0
SIC	15458.9	15266.2

Table 7: AIC and SIC values for the realised GARCH variants

	Symmetric RGARCH	Asymmetric RGARCH
AIC	36374.6	36325.0
SIC	36434.3	365401.7

7.4 Derivations for the asymmetrically decomposed ARGARCH model

Starting from the general form of Realised GARCH described in Hansen et al. (2012),

$$r_t = \mu + \sigma_t \varepsilon_t, \tag{7.3}$$

$$\sigma_{t+1}^2 = \omega + \alpha(r_t - \mu)^2 + \beta\sigma_t^2 +_t, \tag{7.4}$$

$$x_t = \xi + \varphi\sigma_t^2 + \tau(\varepsilon_t) + u_t, \tag{7.5}$$

we look at (7.4) where we are interested in splitting α into α_1 and α_2 for a negative and positive shock in return, respectively. Similarly we decompose γ into γ_1 and γ_2. As in the main paper, we assume the same as in Hansen et al. (2012), most importantly, that ε_t is an i.i.d. normally distributed variable with mean zero and variance 1 and that u_t is i.i.d. with mean 0 and variance σ_u^2. We decompose α and γ using indicator functions for the events of a negative or positive shock. The relation now reads:

$$\sigma_{t+1}^2 = \omega + (\alpha_1 \mathbb{1}_{\{\varepsilon_t<0\}} + \alpha_2 \mathbb{1}_{\{\varepsilon_t\geq 0\}})\sigma_t^2\varepsilon_t^2 + \beta\sigma_t^2 + (\gamma_1 \mathbb{1}_{\{\varepsilon_t<0\}} + \gamma_2 \mathbb{1}_{\{\varepsilon_t\geq 0\}})x_t, \tag{7.6}$$

since from (7.3) it follows that $(r_t - \mu)^2 = \sigma_t^2 \varepsilon_t^2$.

Then in (7.7) we introduce $\frac{\alpha_1+\alpha_2}{2}$ and $\frac{\gamma_1+\gamma_2}{2}$ by using that for any scalar c: $0 = c - c$.

$$\sigma_{t+1}^2 = \omega + \alpha_1 \mathbb{1}_{\{\varepsilon_t<0\}}\sigma_t^2\varepsilon_t^2 + \alpha_2 \mathbb{1}_{\{\varepsilon_t\geq 0\}}\sigma_t^2\varepsilon_t^2 + \frac{\alpha_1+\alpha_2}{2}(\sigma_t^2 - \sigma_t^2) + \beta\sigma_t^2 + \gamma_1 \mathbb{1}_{\{\varepsilon_t<0\}}x_t + \gamma_2 \mathbb{1}_{\{\varepsilon_t\geq 0\}}x_t + \frac{\gamma_1+\gamma_2}{2}(x_t - x_t) \tag{7.7}$$

By rewriting this equation, we obtain

$$\sigma_{t+1}^2 = \omega + \alpha_1\left[\varepsilon_t^2 \mathbb{1}_{\varepsilon_t<0} - \frac{1}{2}\right]\sigma_t^2 + \alpha_2\left[\varepsilon_t^2 \mathbb{1}_{\varepsilon_t\geq 0} - \frac{1}{2}\right]\sigma_t^2 + (\alpha+\beta)\sigma_t^2 + \gamma_1\left[\mathbb{1}_{\varepsilon_t<0} - \frac{1}{2}\right]x_t + \gamma_2\left[\mathbb{1}_{\varepsilon_t\geq 0} - \frac{1}{2}\right]x_t + \gamma x_t, \tag{7.8}$$

where we have defined $\alpha = \frac{\alpha_1+\alpha_2}{2}$ and $\gamma = \frac{\gamma_1+\gamma_2}{2}$. Now, we expand the last term using (7.5) and implement the unconditional expectation of σ_{t+1}^2 of The Realised GARCH model, which reads $\sigma^2 = \frac{\omega+\gamma}{1-\alpha-\beta-\gamma\varphi}$ (Hansen et al., 2012), where now α and γ are defined as mentioned. These implementations give us

$$\sigma_{t+1}^2 = \sigma_t^2 + \alpha_1\left[\varepsilon_t^2 \mathbb{1}_{\varepsilon_t<0} - \frac{1}{2}\right]\sigma_t^2 + \alpha_2\left[\varepsilon_t^2 \mathbb{1}_{\varepsilon_t\geq 0} - \frac{1}{2}\right]\sigma_t^2 + (\alpha+\beta+\gamma\varphi)(\sigma_t^2 - \sigma^2) + \tag{7.9}$$

$$+ \gamma_1\left[\mathbb{1}_{\varepsilon_t<0} - \frac{1}{2}\right]x_t + \gamma_2\left[\mathbb{1}_{\varepsilon_t\geq 0} - \frac{1}{2}\right]x_t. + \gamma(\tau(\varepsilon) + u_t) \tag{7.10}$$

Note that all terms except for σ_t^2 and $(\alpha+\beta+\gamma\varphi)(\sigma_t^2 - \sigma^2)$ have an expected value of 0, regardless of σ_t^2. $\tau(\varepsilon_t)$ and terms involving indicator functions are expected to be zero, as ε_t is normally distributed with mean zero and $\mathbb{E}[\mathbb{1}_{\varepsilon_t<0} - \frac{1}{2}] = \Pr(\varepsilon_t < 0) - \frac{1}{2} = \frac{1}{2} - \frac{1}{2} = 0,$. Moreover, u_t has expected value of zero. This means that for d-day ahead expectation of the daily variance we obtain

$$\mathbb{E}_t[\sigma_{t+d}^2] = \sigma^2 + (\alpha+\beta+\gamma)(\mathbb{E}_t[\sigma_{t+d-1}^2] - \sigma^2), \tag{7.11}$$

which is the same relation as in Hansen et al. (2012) only with a different definition for α and γ. Hence, from here on the derivation for the predicted variance is identical to Hansen et al. (2012), from which we obtain the predicted variance formulated as follows:

$$\mathrm{PV}_{t,d} := \mathbb{E}_t\left[\sum_{i=1}^{d} r_{t+i}^2\right] = d(\mu^2 + \sigma^2) + \frac{1-(\alpha+\beta+\gamma\varphi)^d}{1-\alpha-\beta-\gamma\varphi}(\sigma_{t+1}^2 - \sigma^2) \qquad d \in \{1, 5, 21\} \tag{7.12}$$

but with the important distinction that here $\alpha = \frac{\alpha_1+\alpha_2}{2}$ and $\gamma = \frac{\gamma_1+\gamma_2}{2}$.

7.5 Additional results

Table 8: MSE, QLIKE and $\bar{L}_1$ values for the fixed static forecasts (with minimum values per target variable per loss function bolded)

	Panel A: One day ahead (d = 1)							
	TV1				TV2			
	AGARCH	ARGARCH	VIX	HAR-RV	AGARCH	ARGARCH	VIX	HAR-RV
MSE	20.305	18.124	20.077	**17.497**	6.967	4.437	4.779	**4.326**
QLIKE	1.629	1.557	1.670	**1.515**	0.371	0.312	0.420	**0.267**
$\bar{L}_1$	400.547	379.432	393.058	**366.861**	63.705	39.985	**32.932**	38.119
	Panel B: One week ahead (d = 5)							
	TV1				TV2			
	AGARCH	ARGARCH	VIX	HAR-RV	AGARCH	ARGARCH	VIX	HAR-RV
MSE	287.920	232.925	244.017	**228.516**	157.334	83.342	81.899	**78.357**
QLIKE	0.538	**0.491**	0.529	0.510	0.327	0.295	0.338	**0.283**
$\bar{L}_1$	17715.608	15281.266	15177.716	**14727.514**	6736.338	2986.192	**1524.903**	2467.030
	Panel C: One month ahead (d = 21)							
	TV1				TV2			
	AGARCH	ARGARCH	VIX	HAR-RV	AGARCH	ARGARCH	VIX	HAR-RV
MSE	4552.395	**3183.002**	3203.316	3246.496	3288.459	1602.722	1372.897	**860.517**
QLIKE	0.491	0.494	**0.438**	0.883	**0.378**	0.411	0.473	0.652
$\bar{L}_1$	713869.151	399766.817	**388387.714**	394609.686	541346.522	150558.393	**18057.959**	56525.604