Journal of Banking and Finance 133 (2021) 106285

Contents lists available at ScienceDirect

Journal of Banking and Finance

journal homepage: www.elsevier.com/locate/jbf

A Practical Guide to harnessing the HAR volatility model[☆]

Adam Clements[a,*], Daniel P.A. Preve[b]

[a] School of Economics and Finance, Queensland University of Technology, QUT Business School, GPO Box 2434, Brisbane QLD 4001, Australia
[b] School of Economics, Singapore Management University, 81 Victoria Street, Singapore 188065, Singapore

ARTICLE INFO

Article history:
Received 15 May 2019
Accepted 10 August 2021
Available online 12 August 2021

JEL classification:
C22
C51
C52
C53
C58

Keywords:
Volatility forecasting
Realized variance
HAR
HARQ
Robust regression
Weighted least squares
Box-Cox transformation
Forecast comparisons
QLIKE
MSE
VaR
Model confidence set

ABSTRACT

The standard heterogeneous autoregressive (HAR) model is perhaps the most popular benchmark model for forecasting return volatility. It is often estimated using raw realized variance (RV) and ordinary least squares (OLS). However, given the stylized facts of RV and well-known properties of OLS, this combination should be far from ideal. The aim of this paper is to investigate how the predictive accuracy of the HAR model depends on the choice of estimator, transformation, or combination scheme made by the market practitioner. In an out-of-sample study, covering the S&P 500 index and 26 frequently traded NYSE stocks, it is found that simple remedies systematically outperform not only standard HAR but also state of the art HARQ forecasts.

1. Introduction

Forecasting the volatility of financial asset returns is an important issue in the context of risk management, portfolio construction, and derivative pricing. As such, a great deal of research effort has focused on developing and evaluating volatility forecasting models. With the widespread availability of high-frequency financial data, the recent literature has focused on employing realized volatility (RV) to build forecasting models. The heterogeneous autoregressive (HAR) model of Corsi (2009) was designed to parsimoniously capture the strong persistence typically observed in RV and has become the workhorse of this literature due to its consistently good forecasting performance, and that standard linear regression techniques can be used for its estimation. The influence of this model is reflected in the fact that as of July 2021, Corsi (2009) has attracted more than 2100 citations according to Google Scholar. The original HAR model is often estimated using RV and the method of ordinary least squares (OLS). However, given stylized features of raw RV (such as spikes/outliers, conditional heteroskedasticity, non-Gaussianity) and well-known properties of OLS (highly sensitive to outliers, suboptimal in the presence of conditional heteroskedasticity or non-Gaussianity), this combination should be far from ideal, leaving opportunity for straightforward improvements.

To better deal with the stylized features of RV, a number of straightforward alternatives to using OLS and raw RV for generating forecasts from the HAR model will be considered. First, the impact of alternative estimation schemes, employing weighted least squares (WLS) or robust regression (RR), for the HAR model will be

☆ We thank the editor, Geert Bekaert, and three anonymous referees for comments and suggestions that greatly improved the paper. We also thank Andrew Patton, Jun Yu, Andrey Vasnev, Stephen Thiele, and seminar participants at the University of Sydney for helpful comments and discussions. All remaining errors are our own.

* Corresponding author.
E-mail addresses: a.clements@qut.edu.au (A. Clements), dpreve@smu.edu.sg (D.P.A. Preve).

https://doi.org/10.1016/j.jbankfin.2021.106285
0378-4266/

investigated. Second, the impact of alternative Box-Cox transformations (logarithmic and square root) of RV for the HAR model will be studied. Third, combinations of different estimation and transformation schemes will be explored. The potential benefits of these alternative approaches will be investigated in an out-of-sample study with the HAR model estimated by OLS used as benchmark. For a more complete picture, the recent HARQ model is also used as a benchmark model as it has been documented to outperform not only the original HAR model but also some of its numerous extensions in terms of forecasting. The HARQ model represents the state of the art in volatility forecasting models, and is designed to directly deal with the estimation error in RV (Bollerslev et al., 2016; 2018b). It should be emphasized that, in contrast to Buccheri and Corsi (2019) and Cipollini et al. (2021), the goal here is not to extend the original HAR model but instead to investigate how to get the most out of the existing model. For instance, by carefully selecting its estimator.

The first issue considered is how the predictive accuracy of the HAR model depends on the choice of estimator. The idea of investigating whether the choice of estimator matters for forecasting is not new, and has for instance been considered by Westerlund and Narayan (2012) in the context of stock return predictability. In the context of volatility forecasting, a few estimators have been suggested for the HAR model. While Patton and Sheppard (2015) employed a simple WLS scheme, RR so far does not seem to have been considered. Furthermore, although alternatives to OLS have been proposed, the impact of the choice of estimator on the predictive accuracy of the HAR model does not seem to have been systematically examined. The empirical results reveal that the choice of the estimation scheme under which the HAR parameter estimates are obtained is important for volatility forecasting. By simply considering WLS or RR as an alternative to OLS, considerable reductions in common statistical and economic loss measures are observed. The benefits of replacing OLS with WLS or RR are particularly clear for longer forecast horizons.

Next, how the predictive accuracy of the HAR model depends on the choice of transformation of RV is considered. The idea of using transformed, rather than raw, RV for forecasting is not new and has been considered by Andersen et al. (2003), Corsi (2009), Bekaert and Hoerova (2014) and Taylor (2017) among others. The results of the empirical study show that HAR models based on transformed RV (and estimated by OLS) offer substantial reductions in the loss measures compared to the standard HAR based on raw RV. Overall, the logarithmic transformation does better than the square root transformation. Improvements in forecast accuracy are often similar to those provided by RR.

Finally, the possible benefits of combination schemes, alternative estimation schemes (WLS or RR) applied to HAR models based on transformed (logarithmic or square root) RV, is examined. This idea of combining different transformations with alternative estimators in the context of return volatility predictability appears to be new. It is found that combination schemes often do better than HAR models based on transformed RV, estimated by OLS. A HAR model based on logarithmic RV estimated using RR does best overall. As the loss functions used for forecast comparison deal with over- and under-prediction differently, the interaction between forecasts from the alternative approaches and the loss functions is further investigated and discussed.

The results are robust to index and individual stock data, with or without an "insanity filter" applied to all forecasts, to periods of high and low volatility, and to alternative multi-step ahead forecast schemes (direct, indirect). The approaches considered here also perform well compared to some extended HAR models (LHAR, HAR-RSV, HAR-CJ) suggested in the literature. These findings provide useful practical insights in the application of the HAR model. Improvements in forecast accuracy can readily be obtained without the need to resort to data beyond publicly available RV and sophisticated extensions of the HAR model.

The remainder of this paper is organized as follows. Section 2 describes its methodology, Section 3 reports the results of its empirical study, and Section 4 concludes. Additional results to complement the main paper are contained in an Online Appendix.

2. Methodology

This section describes the measures and models used to construct volatility forecasts, the estimators and transformations employed, how forecasts are computed, and how their accuracy is assessed.

2.1. Realized variance

We consider a single asset for which the log-price process P within the active part of a trading day evolves in continuous time as

$$dP_t = \mu_t dt + \sigma_t dW_t, \tag{1}$$

where μ and σ are the instantaneous drift and volatility processes, respectively, and W is a standard Brownian motion (Wiener process). The ith Δ-period return within day t is defined as

$$r_{t,i} = P_{t-1+i\Delta} - P_{t-1+(i-1)\Delta}, \quad i = 1, 2, \dots, M,$$

where $M = 1/\Delta$ is the sampling frequency. Hence, the daily logarithmic return for the active part of trading day t is $r_t = \sum_{i=1}^{M} r_{t,i}$.

In the simplest case, we wish to forecast the latent one-day integrated variance defined by

$$IV_t = \int_{t-1}^{t} \sigma_s^2 ds. \tag{2}$$

Although (2) is unobservable it can be consistently estimated by the one-day realized variance (RV)

$$RV_t = \sum_{i=1}^{M} r_{t,i}^2,$$

as $M \to \infty$ (Andersen and Bollerslev, 1998). Hence, the RV measure is defined as the sum of the squared returns within day t. Given restrictions on the sampling frequency M, Barndorff-Nielsen and Shephard (2002) show that the estimation error in RV can be characterized by $RV_t = IV_t + \eta_t$, where

$$\frac{\eta_t}{\sqrt{2\Delta IQ_t}} = \frac{RV_t - IV_t}{\sqrt{2\Delta IQ_t}}$$

is approximately $N(0,1)$, or standard normal, and $IQ_t = \int_{t-1}^{t} \sigma_s^4 ds$ is the integrated quarticity (IQ) which can be consistently estimated by the realized quarticity (RQ)

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

Analogous results are available for common transformations of RV. See, for example, Corsi et al. (2008). For square root transformed RV,

$$\frac{\sqrt{RV_t} - \sqrt{IV_t}}{\sqrt{\frac{\Delta}{2} \frac{RQ_t}{RV_t}}} \tag{4}$$

is approximately $N(0,1)$. For log transformed RV,

$$\frac{\log RV_t - \log IV_t}{\sqrt{2\Delta \frac{RQ_t}{(RV_t)^2}}} \tag{5}$$

is approximately $N(0,1)$, where log denotes the natural logarithm.

2.2. The HAR & HARQ models

2.2.1. HAR

With the widespread availability of high-frequency intraday data, the recent literature has focused on employing RV to build forecasting models for time-varying return volatility. Among these forecasting models, the HAR model proposed by Corsi (2009) has gained popularity due to its simplicity and consistent forecasting performance in applications. The formulation of the HAR model is based on a straightforward extension of the so-called heterogeneous ARCH, or HARCH, class of models analyzed by Muller et al. (1997). Under this approach, the conditional variance of the discretely sampled returns is parameterized as a linear function of lagged squared returns over the same horizon together with the squared returns over longer and/or shorter horizons.

The original HAR model specifies RV as a linear function of daily, weekly and monthly realized variance components, and can be expressed as

$$RV_t = \beta_0 + \beta_1 RV^d_{t-1} + \beta_2 RV^w_{t-1} + \beta_3 RV^m_{t-1} + u_t, \tag{6}$$

where the β_j ($j = 0, 1, 2, 3$) are unknown parameters that need to be estimated, RV_t is the realized variance of day t, and $RV^d_{t-1} = RV_{t-1}$, $RV^w_{t-1} = \frac{1}{5}\sum_{i=1}^{5} RV_{t-i}$, $RV^m_{t-1} = \frac{1}{22}\sum_{i=1}^{22} RV_{t-i}$ denote the daily, weekly and monthly lagged realized variance, respectively. This specification of RV parsimoniously captures the high persistence observed in most realized variance series.

2.2.2. HARQ

Bollerslev et al. (2016) recently proposed an easily implemented, and by OLS estimated, extension of the HAR model dubbed the HARQ model, which accounts for the error with which RV is estimated by using RQ. The full HARQ (HARQ-F) model can be written as

$$\begin{aligned} RV_t = {} & \beta_0 + (\beta_1 + \beta_{1Q}\sqrt{RQ^d_{t-1}})RV^d_{t-1} + (\beta_2 + \beta_{2Q}\sqrt{RQ^w_{t-1}})RV^w_{t-1} \\ & + (\beta_3 + \beta_{3Q}\sqrt{RQ^m_{t-1}})RV^m_{t-1} + u_t, \end{aligned}$$

where (similar to the original HAR model) RQ^w_{t-1}, RQ^w_{t-1} and RQ^m_{t-1} denote the daily, weekly, and monthly lagged realized quarticity, respectively. Bollerslev et al. (2016) find that, at least for short-term forecasting, a simplified version

$$RV_t = \beta_0 + (\beta_1 + \beta_{1Q}\sqrt{RQ^d_{t-1}})RV^d_{t-1} + \beta_2 RV^w_{t-1} + \beta_3 RV^m_{t-1} + u_t, \tag{7}$$

is useful as most of the attenuation bias in the forecasts (due to RV being less persistent than unobserved IV) is due to the estimation error in RV^d_{t-1}. Overall, this framework allows for less weight to be placed on historical observations of RV when the measurement error captured by RQ is higher.

The subsequent study considers the forecasting performance of the original HAR model, when its parameters are estimated using alternative methods to OLS, and when it is fitted to transformed rather than raw RV. The standard HAR model (6) and its (state of the art) HARQ extension (7), both estimated using OLS, are then used as benchmarks models.

2.3. The estimators

The HAR model in (6) is often estimated using RV and the method of OLS. However, given the stylized facts of RV (such as spikes/outliers, conditional heteroskedasticity, and non-Gaussianity) and well-known properties of OLS, this combination should be far from ideal. Instead alternative methods such as weighted least squares (WLS) and robust regression (RR) seem more appropriate. Next we briefly review the above methods, and the associated estimation schemes used in our out-of-sample forecasting study.

2.3.1. OLS

For the HAR model, the OLS estimator of $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2, \beta_3)$ given the observations $RV_1, \ldots, RV_n$ is the solution to the minimization problem

$$\min_{b_0,b_1,b_2,b_3} \sum_{t=23}^{n} (RV_t - b_0 - b_1 RV^d_{t-1} - b_2 RV^w_{t-1} - b_3 RV^m_{t-1})^2.$$

It is well-known that if the errors u_t in autoregressions such as (6) are independent, normally (Gaussian) distributed, and homoskedastic the optimal (in a asymptotic efficiency sense) estimator of $\boldsymbol{\beta}$ is the OLS estimator.

2.3.2. WLS

Weighted least squares attempts to provide a more efficient alternative to OLS. Instead of the sum of squared deviations, their weighted sum is minimized. For the HAR model, the WLS estimator of $\boldsymbol{\beta}$ is the solution to the minimization problem

$$\min_{b_0,b_1,b_2,b_3} \sum_{t=23}^{n} w_t (RV_t - b_0 - b_1 RV^d_{t-1} - b_2 RV^w_{t-1} - b_3 RV^m_{t-1})^2, \tag{8}$$

where $w_t > 0$ is the weight of the tth observation. If each weight w_t is inversely proportional to the conditional variance of the corresponding error u_t, then the WLS estimator is more efficient than the OLS estimator. In this way, less weight is given to errors which are likely to be large. While Patton and Sheppard (2015) employed a simple WLS scheme (described below) for estimating the HAR model, they never considered alternatives. Here we will examine this method more thoroughly by introducing a number of different WLS schemes.

The first scheme uses RQ to directly capture the heteroskedasticity in RV, defining w_t in (8) as $w_t = 1/\sqrt{RQ_{t-1}}$. This approach is closest in spirit to the HARQ model (7), and places less weight during estimation on periods where volatility is less precisely estimated.[1] This scheme will hereafter be denoted by WLS_{RQ}-HAR. Given the strong positive empirical correlation between RV and $\sqrt{\text{RQ}}$, the second scheme uses RV directly with weights $w_t = 1/RV_{t-1}$. This specification is useful for the many instances where RV is publicly available but RQ is not. This scheme is denoted WLS_{RV}-HAR. The third WLS scheme is the approach briefly discussed in Patton and Sheppard (2015). This scheme, denoted $\text{WLS}_{\widehat{\text{RV}}}$-HAR, uses weights $w_t = 1/\widehat{RV}_t$, where $\widehat{RV}_t$ is the fitted value from the standard HAR model (6) estimated using OLS. Given the positive relationship between volatility and RQ, this scheme places less weight during estimation on periods where volatility is less precisely estimated without requiring RQ directly. Corsi et al. (2008) analyse the residuals of HAR models estimated by OLS and find evidence of conditional heteroskedasticity, which motivates the authors to consider HAR-GARCH specifications. Influenced by their findings, the fourth scheme considers a three-step estimation approach for the HAR model: The first step is to estimate its parameters using OLS and compute residuals. The second step is to estimate a GARCH(1,1) on the OLS residuals. The third step is to use these estimates to fit the HAR model by WLS with weights $w_t = 1/\hat{h}_t$, where $\hat{h}_t$ is the fitted value of the conditional

[1] Alternatively, $w_t = 1/\sqrt{RQ_t}$ could be used here, as the weights are used only for parameter estimation and not for forecasting. In practice, however, using RQ_t instead of RQ_{t-1} makes very little difference to estimation. Instead, it is the time variation in the weights reflecting the heteroskedasticity that is important. Given the persistence in RQ, there is no practical difference between the relative levels through time of either RQ_t or RQ_{t-1}.

variance of the GARCH(1,1). The final step is partially motivated by Romano and Wolf (2017) who find that WLS can be superior to OLS even when the model used to estimate the heteroskedastic function is misspecified. The weighting scheme outlined above is denoted WLS_{G}-HAR. As the previous two estimation schemes, this final scheme has the practical benefit of employing WLS without resorting to RQ.

Although four different WLS schemes have been described above, only results for the WLS_{RQ}-HAR and WLS_{G}-HAR will be reported in the empirical study. This is mainly for the sake of brevity as quite a number of estimation, transformation and combination schemes, as well as benchmarks, will be considered. Because of the strong correlation between RV and $\sqrt{\text{RQ}}$, the performance of the WLS_{RV}-HAR is very similar to that of the WLS_{RQ}-HAR. Accordingly, results for the WLS_{RV}-HAR will not be reported. Unreported results show that while the $\text{WLS}_{\widehat{\text{RV}}}$-HAR generally does better than the OLS-HAR, its forecasting performance typically is somewhat behind that of the WLS_{RQ}-HAR and WLS_{G}-HAR. Consequently, the $\text{WLS}_{\widehat{\text{RV}}}$-HAR will not be used in the empirical study.

2.3.3. RR

Although optimal under ideal conditions, the OLS estimator is also well-known to be highly sensitive to outliers (unusual observations) in the data. For this reason more robust estimators, such as the commonly used *M*-estimator, have been proposed as alternatives. For the HAR model, the *M*-estimator of $\boldsymbol{\beta}$ is the solution to the minimization problem

$$\min_{b_0,b_1,b_2,b_3} \sum_{t=23}^{n} \rho(RV_t - b_0 - b_1 RV_{t-1}^d - b_2 RV_{t-1}^w - b_3 RV_{t-1}^m),$$

where ρ is a prespecified symmetric function with a unique minimum at zero. An important special case is the least absolute deviations (LAD) estimator, with loss function $\rho(e) = |e|$. For this case the sum of absolute instead of squared deviations is minimized. Hence, by comparison, OLS gives more weight to large deviations (outliers) than LAD.

For many cases, the robust *M*-estimate of $\boldsymbol{\beta}$ is computed using iteratively reweighted least squares (IRLS) with the weight function $w(e) = \rho'(e)/e$, where ρ' is the derivative of ρ. In our empirical study, we use Tukey's bisquare (or biweight) estimator for which

$$\rho(e) = \begin{cases} \frac{k^2}{6}\left\{1 - \left[1 - \left(\frac{e}{k}\right)^2\right]^3\right\}, & |e| \le k \\ \frac{k^2}{6}, & |e| > k \end{cases} \quad \text{and}$$

$$w(e) = \begin{cases} \left[1 - \left(\frac{e}{k}\right)^2\right]^2, & |e| \le k \\ 0, & |e| > k \end{cases}$$

where k is the tuning constant.[2] This approach will hereafter be denoted by RR-HAR. The popular bisquare and LAD *M*-estimators are considered resistant, or robust. That is, they are relatively unaffected by outliers.

2.4. The transformations

An alternative to employing estimation methods other than OLS is to use transformations. The logarithmic transformation, for example, is known to be appropriate for series whose standard deviation increases linearly with the mean (Brockwell and Davis, 1991). Numerous alternative transformations have been proposed. The best known perhaps being the Box-Cox transformations (Box and Cox, 1964), which is a family of variance-stabilizing transformations. Transformations belonging to this family are often used in practice to obtain a model with a simple structure, and (close to) normally distributed errors with constant variance. The Box-Cox transformation of a time series variable y_t is

$$y_t(\lambda) = \begin{cases} \frac{y_t^\lambda - 1}{\lambda}, & \lambda \ne 0, \\ \log y_t, & \lambda = 0, \end{cases}$$

where λ is the power parameter. In the context of modeling and forecasting RV, important special cases include the logarithmic transformation ($\lambda = 0$), the quartic root transformation ($\lambda = 1/4$), and the square root transformation ($\lambda = 1/2$). See Corsi (2009), Taylor (2017), and the references therein.

To highlight the impact of such transformations, Fig. 1 shows the distribution of raw RV for the S&P 500 series used in Bollerslev et al. (2016), along with the distributions of sqr-, qr- and log-transformed RV. The top panel illustrates well-known features of the RV, which is nonnegative with a distribution exhibiting substantial skewness and excess kurtosis. It is clear from the lower panel that all three transformations result in more symmetric, approximately Gaussian, distributions. The sample skewness of raw RV exceeds 10, while the skewness of the sqr-transformed RV is 3.2, 1.5 for the qr-transformed data, and 0.5 for the log-transformed data. In sum, all three transformations appear useful for reducing skewness, and hence the possible effect of outliers and potential heteroskedasticity in the RV series. Nonetheless, for the sake of brevity, only log- and sqr-transformed RV will be used in the empirical study.

2.5. Combinations

A natural question that arises in this context is if there are benefits to combination schemes, that is, to alternative estimation schemes applied to HAR models based on transformed RV. Here the RR and two WLS schemes will be used to examine how the predictive accuracy of HAR models based on logarithmic or square root transformed RV depends on the choice of estimator. For logarithmic RV, the resulting combination schemes are denoted RR-log-HAR, WLS_{RQ}-log-HAR and WLS_{G}-log-HAR. Similarly, for square root transformed RV, the schemes are denoted RR-sqr-HAR, WLS_{RQ}-sqr-HAR and WLS_{G}-sqr-HAR. The weights for WLS_{RQ}-sqr-HAR, $w_t = \sqrt{RV_{t-1}/RQ_{t-1}}$, are based on (4), and the WLS_{RQ}-log-HAR weights, $w_t = RV_{t-1}/\sqrt{RQ_{t-1}}$, are based on (5). The purpose of the above six combination schemes is to investigate if the alternative estimators can correct for possible shortcomings of the transformations, such as remaining heteroskedasticity in the transformed RV.

2.6. Forecasting

2.6.1. Raw RV

The optimal (in the MSE sense) forecast of RV_t for the HAR model (6) given the information set at time $t-1$ can be expressed as

$$F_t = \beta_0 + \beta_1 RV_{t-1} + \frac{\beta_2}{5}\sum_{i=1}^{5} RV_{t-i} + \frac{\beta_3}{22}\sum_{i=1}^{22} RV_{t-i}.$$

Similarly, for the HARQ model (7)

$$F_t = \beta_0 + (\beta_1 + \beta_{1Q}\sqrt{RQ_{t-1}^d})RV_{t-1} + \frac{\beta_2}{5}\sum_{i=1}^{5} RV_{t-i} + \frac{\beta_3}{22}\sum_{i=1}^{22} RV_{t-i}.$$

Following Bollerslev et al. (2016), weekly or monthly *direct* forecasts are obtained by replacing the daily RVs on the left-hand-sides of (6) and (7) with the weekly or monthly RVs.

[2] We use MATLAB's `robustfit` function, with its default tuning constant, to compute the bisquare estimates. We also tried the LAD estimator, but were not able to improve upon our results reported for the bisquare estimator.

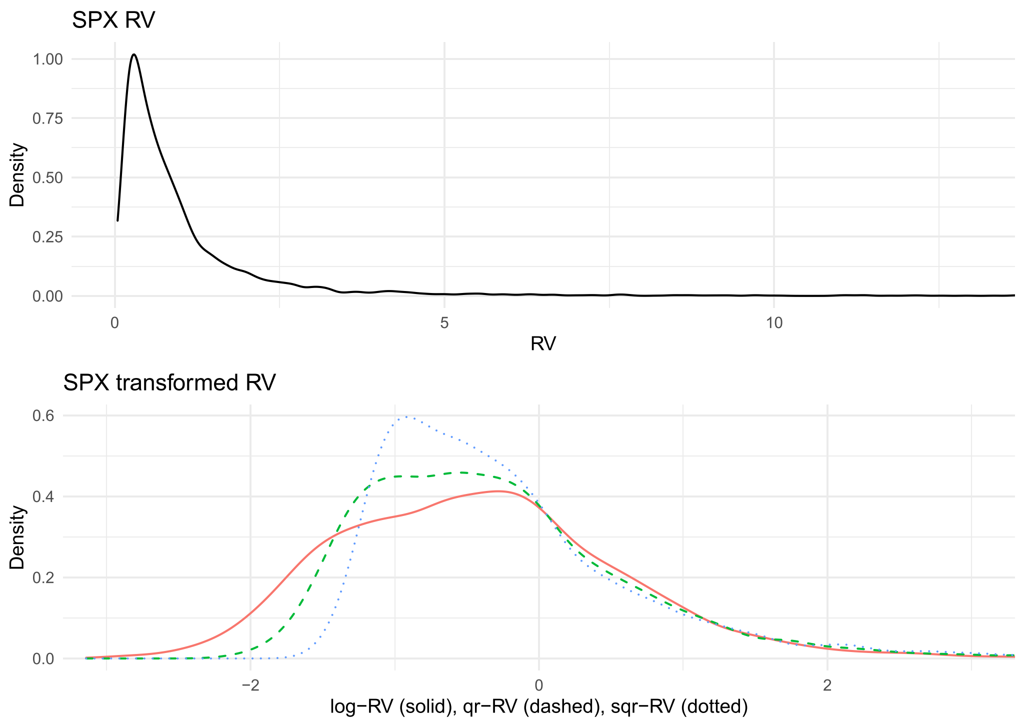

Fig. 1. Top panel: Kernel density estimate of the S&P 500 RV observations used in Section 3.2. Bottom panel: Kernel density estimates of the log-RV observations (solid), qr-RV observations (dashed), and sqr-RV observations (dotted).

Table 1
In-sample daily model estimates and measures of fit for some selected (RR, WLS_{RQ}, WLS_G) estimation schemes. Results based on the full SPX sample. The benchmark HAR and HARQ models are estimated by OLS. Robust standard errors are reported in parentheses, together with the R^2 and empirical MSE and QLIKE.

	HAR		HARQ		RR-HAR		WLS_{RQ}-HAR		WLS_G-HAR	
β_0	0.1126	(0.0615)	-0.0099	(0.0617)	0.1126	(0.0057)	0.0517	(0.0117)	0.0223	(0.0113)
β_1	0.2273	(0.1104)	0.5929	(0.0838)	0.3713	(0.0035)	0.5781	(0.0253)	0.4310	(0.0317)
β_2	0.4904	(0.1352)	0.3586	(0.1284)	0.2257	(0.0059)	0.2391	(0.0277)	0.4758	(0.0466)
β_3	0.1864	(0.1010)	0.0976	(0.1052)	0.1165	(0.0052)	0.1548	(0.0225)	0.0972	(0.0314)
β_{1Q}			-0.3602	(0.0637)						
R^2	0.5224		0.5624		0.4933		0.4773		0.4944	
MSE	2.5728		2.3482		2.7802		2.8163		2.7254	
QLIKE	0.1439		0.1358		0.1512		0.1340		0.1331	

2.6.2. Box-Cox transformed RV

From Table 1 in Proietti and Lütkepohl (2013), a bias-corrected forecast of RV_t for the HAR model applied to *logarithmic* (instead of raw) daily RV is

$$F_t = \exp\left\{\beta_0 + \beta_1 \log RV_{t-1} + \frac{\beta_2}{5}\sum_{i=1}^{5}\log RV_{t-i} + \frac{\beta_3}{22}\sum_{i=1}^{22}\log RV_{t-i} + \frac{\sigma_t^2}{2}\right\}, \tag{9}$$

where σ_t^2 is the conditional variance of the errors u_t. Moreover, a bias-corrected forecast of RV_t for the same model applied to *square root* daily RV is

$$F_t = N_t\left(1 + \frac{1}{4}\frac{\sigma_t^2}{N_t}\right), \tag{10}$$

where N_t denotes the naïve forecast,

$$N_t = \left[1 + \frac{\beta_0}{2} + \beta_1\left(\sqrt{RV_{t-1}} - 1\right) + \frac{\beta_2}{5}\sum_{i=1}^{5}\left(\sqrt{RV_{t-i}} - 1\right) + \frac{\beta_3}{22}\sum_{i=1}^{22}\left(\sqrt{RV_{t-i}} - 1\right)\right]^2.$$

The forecasts in (9) and (10) are optimal if the transformed series is normally distributed. For the sake of simplicity, we use the sample variance of the residuals from the estimation window to estimate the variance terms in (9) and (10).

2.6.3. Insanity filter

The HARQ model may on rare occasions generate implausibly large or small forecasts. For this reason, Bollerslev et al. (2016) follow Swanson and White (1997) in applying an "insanity filter" (IF) to all forecasts. Different IFs have been suggested in the related literature (Patton and Sheppard, 2015; Bollerslev et al., 2018a). Here the IF specification of Bollerslev et al. (2016) will be used. With this filter, any forecast greater than the maximum, or less than the minimum, of the dependent variable observed in the estimation period is replaced by the sample average over that period. That is, "insanity" is replaced by "ignorance".

2.7. Comparing forecast accuracy

2.7.1. Loss measures

Following the literature on volatility forecast comparison (Patton, 2011; Patton and Sheppard, 2009), the empirical quasi-likelihood (QLIKE) and mean squared error (MSE) will be used to assess out-of-sample forecast accuracy. For daily RV, these statistical measures are defined as

$$\text{QLIKE} = \frac{1}{T}\sum_{t=1}^{T}\left(\frac{RV_t}{F_t} - \log\frac{RV_t}{F_t} - 1\right), \tag{11}$$

and

$$\text{MSE} = \frac{1}{T}\sum_{t=1}^{T}(RV_t - F_t)^2, \tag{12}$$

where T is the number of forecasts and F_t denotes a forecast of RV_t (which proxies for IV_t) from the different models or approaches.[3] Eqs. (11) and (12) are easily modified for weekly, or longer horizon volatility forecasts.

Bee et al. (2016) among others use HAR models to forecast risk estimates such as Value-at-Risk (VaR), a critically important issue in the context of prudential regulation. Here, forecasts will also be compared under an economic loss function directly related to VaR. The conditional VaR is given by

$$\text{VaR}_t^{\alpha} = \mu_t + \Phi^{-1}(\alpha)\sqrt{F_t} \tag{13}$$

where μ_t is the conditional mean of the return (which for simplicity here is taken to be time-invariant), and Φ^{-1} is the inverse of the standard normal cdf. A quantile estimation based loss function that penalizes VaR violations more heavily will be used to evaluate all forecasts. Specifically, the smoothed loss function proposed by González-Rivera et al. (2004) is used,

$$\widetilde{Q} = \frac{1}{T}\sum_{t=1}^{T}\left[\alpha - \frac{1}{1+e^{\delta(r_t - \text{VaR}_t^{\alpha})}}\right](r_t - \text{VaR}_t^{\alpha}), \tag{14}$$

where the parameter $\delta > 0$ controls the smoothness. Lower values of $\widetilde{Q}$ indicate more accurate forecasts. Following Fameliti and Skintzi (2020), values of $\delta = 25$ and $\alpha = 0.05$ are used in the subsequent empirical analysis. For ease of interpretation, the VaR-based loss measure (14) will hereafter be denoted by VaR.

This paper follows the standard practice of estimating model parameters with respect to loss functions that differ from the loss functions used to assess forecast performance, that is, from the intended use of the model forecasts. Hansen and Dumitrescu (2018) show that likelihood-based estimation is preferred over estimation under the intended loss function whenever the likelihood is correctly specified. However, when the likelihood is misspecified, it is not clear if this remains to be the case. Here, OLS amounts to likelihood-based estimation under the assumption of normality, which is far from the case for RV. Hence, further improvements in predictive accuracy could be made by using the same loss function that is used to assess out-of-sample performance for parameter estimation. A thorough examination of the conflict between the loss function used for estimating model parameters and that used to assess forecast performance is beyond the scope of this paper. Here, a range of well-known estimation methods (OLS, WLS, RR) are used to obtain HAR model parameter estimates. These different approaches are then compared under a range of widely-used loss functions (QLIKE, MSE, VaR).

[3] Simulation based evidence by Patton and Sheppard (2009) suggests the use of QLIKE rather than MSE due to the superior power of QLIKE in Diebold and Mariano (1995) and West (1996) type tests for equal predictive accuracy (EPA).

2.7.2. The model confidence set

Statistically significant differences in forecast performance will be assessed using the model confidence set (MCS) introduced by Hansen et al. (2011).[4] The MCS procedure avoids the specification of a benchmark model, and starts with a collection of competing models (or approaches), $\mathcal{M}^0$, indexed by $i = 1, \ldots, m_0$. For each of the loss functions in Section 2.7.1, loss differentials $d_{ij,t}$ between models i and j are computed, and $H_0 : E(d_{ij,t}) = 0$ for all i, j (the null hypothesis of EPA) is tested. If the null hypothesis is rejected at the significance level α, the worst performing model is eliminated and the process is repeated until non-rejection occurs with the set of surviving models being the MCS, $\widehat{\mathcal{M}}^*_{1-\alpha}$. By using the same significance level for all tests, $\widehat{\mathcal{M}}^*_{1-\alpha}$ contains the best model(s) from $\mathcal{M}^0$ with a limiting $(1-\alpha)$ level of confidence.[5] Here the tests for EPA employ the range statistic described in Hansen et al. (2003). Since the MCS should be used with caution when forecasts are based on estimated parameters and models are nested (Hansen et al., 2011), a 90% MCS will be complemented with loss ratios of the standard HAR to alternative approaches.

3. Empirical results

3.1. Data

The empirical study here is based on the Standard & Poor's 500 (SPX) index and 26 frequently traded NYSE stocks. For the S&P 500, the same series of RV and RQ used in Bollerslev et al. (2016) are employed.[6] This dataset spans 21 April 1997 to 30 August 2013 representing 4096 daily observations and was chosen as the HARQ model is one of the benchmarks considered here, and as it was central to the original work of Bollerslev et al. (2016). In addition to the SPX, 26 individual stocks (all constituents of the Dow Jones index) are used. Estimates of RV_t and RQ_t are also here based on 5 minute intraday returns downloaded from Thomson Reuters Datascope (new version of Tick History). The data spans 22 July 2002 to 27 June 2019 representing 4193 trading days. A list of ticker symbols and company names for the stocks is provided in the Online Appendix.

3.2. In-sample results

While the practical use of the HAR model in terms of forecasting is the focus of this paper, this section outlines various in-sample estimation results to provide some insights into the important features of the estimation and transformation schemes. The results are for the full SPX sample, and a 1-day forecast horizon, so that the HAR and HARQ in-sample estimation results of Bollerslev et al. (2016) can be used a benchmark for comparative purposes.

Table 1 presents the in-sample estimation results for the RR, WLS_{RQ}, and WLS_{G} estimation schemes. The benchmark HAR and HARQ models, both estimated by OLS, reproduce the results of Bollerslev et al. (2016). Bollerslev et al. (2016) find that the HARQ model generally places greater weight on the daily lag, and lesser weight on the weekly and monthly lags, compared to the HAR model. This is reflected in the HAR and HARQ estimation results in Table 1, with an increase in β_1 and decrease in β_2 and β_3 estimates moving from the HAR to the HARQ model. By design,

[4] The MCS results presented here were obtained using the `mcs` function from the Oxford MFE Toolbox developed by Kevin Sheppard, https://www.kevinsheppard.com/code/matlab/mfe-toolbox/.

[5] In this sense, the MCS at level α is similar to a $(1-\alpha)$% confidence interval for an unknown parameter.

[6] The two series were obtained from Andrew Patton's research page, http://public.econ.duke.edu/~ap172/research.html.

Table 2
In-sample daily model estimates and measures of fit for some selected (log, sqr) transformation schemes. Results based on the full SPX sample. The benchmark HAR and HARQ models are estimated by OLS on raw RV. Robust standard errors are reported in parentheses, together with the R^2 and empirical MSE and QLIKE.

	HAR		HARQ		log-HAR		sqr-HAR	
β_0	0.1126	(0.0615)	-0.0099	(0.0617)	-0.0204	(0.0089)	-0.0092	(0.0010)
β_1	0.2273	(0.1104)	0.5929	(0.0838)	0.3924	(0.0217)	0.3968	(0.0183)
β_2	0.4904	(0.1352)	0.3586	(0.1284)	0.4082	(0.0321)	0.3857	(0.0284)
β_3	0.1864	(0.1010)	0.0976	(0.1052)	0.1531	(0.0242)	0.1616	(0.0231)
β_{1Q}			-0.3602	(0.0637)				
R^2	0.5224		0.5624		0.5362		0.5268	
MSE	2.5728		2.3482		2.4994		2.5500	
QLIKE	0.1439		0.1358		0.1336		0.1437	

the (OLS-)HARQ with its four predictors has the highest in-sample R^2 and lowest in-sample MSE, followed by the (OLS-)HAR. Results in the subsequent columns show that the choice of estimation scheme has a considerable impact on the estimates for the β_0 through β_3 parameters of the HAR model. *M*-estimation (RR) has a similar effect to HARQ regression, with an increase in β_1 and corresponding decreases in β_2 and β_3 estimates relative to the HAR model. In contrast to the HARQ model, this comes at the expense of in-sample fit relative to the HAR model. The WLS schemes have a similar impact on the weighting of past RV, placing greater weight on the daily lag, and lesser weight on the weekly and monthly lags, compared to the HAR model. WLS_{RQ}-HAR produces a weighting of past RV that is relatively similar to HARQ, with an increase in β_1 and decrease in β_2 and β_3 estimates relative to the HAR. While the WLS_{G}-HAR also places greater weight on the daily lag relative to the HAR, the effect is not as strong as for the HARQ or WLS_{RQ}-HAR. In contrast to OLS, WLS (and RR) does not by design minimize in-sample MSE. Therefore it is not surprising to see inferior R^2 and MSE values for the WLS schemes. Superior in-sample QLIKE values are, however, observed for both WLS schemes compared to the HAR(Q). While the in-sample results for the alternative estimation schemes are mixed, their practical benefit will become clear in the subsequent out-of-sample forecasting exercise.

While not the focus here, a brief discussion of how WLS deals with the heteroskedasticity in RV is still worthwhile. To economise on space, plots of the weights used for the four different schemes in Section 2.3.2 are presented in the Online Appendix. Each of these WLS schemes are based on the reciprocal of a quantity related to financial volatility. Consequently, weights are smaller during periods of market stress and high volatility in 2008–2009, and larger during periods of low volatility. While WLS is clearly a different approach, it is similar in spirit to the HARQ. The HARQ model directly adjusts the parameter on daily lagged RV in proportion to the magnitude of $\sqrt{RQ}$. Instead of allowing parameters to vary as a function of $\sqrt{RQ}$, WLS_{RQ} places less weight on days with high $\sqrt{RQ}$ during estimation. The other three weighting schemes work in a similar fashion. The $w_t = 1/RV_{t-1}$ weights for the WLS_{RV} scheme closely resemble those for WLS_{RQ}. This is not surprising given the strong positive correlation observed between RV_t and $\sqrt{RQ_t}$. While displaying the same general pattern, the weights for the WLS_{G} and $\text{WLS}_{\widehat{\text{RV}}}$ schemes are more muted. The similarity between the WLS_{RQ}, WLS_{RV} and WLS_{G} weights has an important practical implication. As RQ estimates are not always publicly available, and as IQ is notoriously difficult to estimate in finite samples, the two latter schemes offer viable alternatives for dealing with the heteroskedasticity in RV using WLS.

Table 2 reports the in-sample estimation results for the logarithmic ($\lambda = 0$) and square root ($\lambda = 1/2$) transformation schemes. The HAR and HARQ results are reproduced here for reference. While it is not meaningful to directly compare HAR parameter estimates for transformed RV to those obtained for raw RV, comparing estimates for a specific transformation is. Here the (OLS-)log-HAR places greater weight on more recent RV, with its first two AR parameter estimates being considerably larger than the last. The AR parameter estimates for the sqr-HAR model are similar to those for the log-HAR model. Given the similarities observed in Fig. 1 this is unsurprising. However, in contrast to the log-HAR, the sqr-HAR offers only small improvements in in-sample measures of fit over the HAR. The measures were calculated using Eqs. (9) and (10), respectively.

Finally, Table 3 presents in-sample estimation results for two combination schemes to highlight any benefits of employing alternative estimation schemes, here WLS_{RQ}, to transformed RV. The log- and sqr-HAR results are reproduced for ease of comparison. While only a selection of all possible combination schemes are presented here for the sake of brevity (results for RR-log and WLS_{G}-sqr are, for example, not reported), the results highlight the general outcome for the combination schemes. The first two columns of Table 3 show that there are only minor changes in the estimates of the AR parameters of the log-HAR model after WLS (WLS_{RQ}-log-HAR) is used in place of OLS for estimation. Moreover, there is virtually no change in the in-sample measures of fit. The results for square root transformed RV in the last two columns (sqr-HAR and WLS_{RQ}-sqr-HAR) are similar. Overall, these results suggest that there is little impact from employing transformation schemes to RV in combination with alternative HAR estimation schemes. Either a transformation scheme or an alternative estimation scheme is impactful, but not both. Of course, the possible benefits of combination schemes in the context of volatility prediction is the main focus here, an issue which will be addressed below.

While none of the approaches considered here are computationally burdensome, there are (very) small differences in computational cost between schemes to note. More details relating to this issue are provided in the Online Appendix. OLS estimation is, of course, fastest to compute a one-step-ahead HAR forecast. The (non-GARCH) WLS schemes are only marginally slower, followed by the transformation schemes with their more complicated forecast expressions. Due to the iterative nature of IRLS, the RR-HAR approach is slower again but still takes only a fraction of a second to estimate the unknown HAR parameters and generate a forecast. Finally, the WLS_{G}-HAR is the least computationally efficient approach due to its associated numerical optimization. While it is clear that there are differences in computational cost, the actual computational times for all of the schemes are negligible.

The results of this section showed that many of the approaches considered have a similar effect on HAR parameter estimates, and provided some new insights into the impacts of the alternative estimation and transformation schemes. While the results for in-sample fit were mixed, the important practical benefits (as there

Table 3
In-sample daily model estimates and measures of fit for some selected combinations of transformation and estimation schemes. Results based on the full SPX sample. The log-HAR and sqr-HAR models are estimated by OLS. Robust standard errors are reported in parentheses, together with the R^2 and empirical MSE and QLIKE.

	log-HAR		WLS_{RQ}-log-HAR		sqr-HAR		WLS_{RQ}-sqr-HAR	
β_0	-0.0204	(0.0089)	-0.0112	(0.0085)	-0.0092	(0.0010)	0.0025	(0.0089)
β_1	0.3924	(0.0217)	0.4149	(0.0187)	0.3968	(0.0183)	0.4685	(0.0203)
β_2	0.4082	(0.0321)	0.3835	(0.0284)	0.3857	(0.0284)	0.3252	(0.0280)
β_3	0.1531	(0.0242)	0.1569	(0.0221)	0.1616	(0.0231)	0.1619	(0.0222)
R^2	0.5362		0.5365		0.5268		0.5213	
MSE	2.4994		2.4976		2.5500		2.5796	
QLIKE	0.1336		0.1335		0.1437		0.1433	

are no significant costs in terms of computation) of the approaches will be shown in the following out-of-sample forecasting exercise.

3.3. Out-of-sample results

It is well documented that volatility exhibits changes in levels over time. See, for instance, Cai (1994) and Engle and Rangel (2008). This implies instability in linear models of RV, which may be dealt with by introducing time-varying parameters. A successful recent example of this approach in the context of the HAR model is Buccheri and Corsi (2019). As mentioned earlier, the goal of this paper is not to extend the original HAR model but rather to investigate how to get the most out of it. To this end, a rolling forecasting scheme is relatively attractive when one wishes to deal with parameter drift that is difficult to model explicitly (West, 2006). The out-of-sample results reported here in the main paper are based on a rolling scheme with a 1000 day rolling window as in Bollerslev et al. (2016), and Taylor (2017). Results based on a recursive (increasing window) forecasting scheme are consistent with those reported here and are available in the Online Appendix.

Since the original HAR model will be estimated using a rolling scheme, with a limited number of observations, changes in its parameter estimates are to be expected over time. Before discussing the out-of-sample forecasting results, it is instructive to investigate these changes for some selected estimation schemes. Fig. 2 shows trajectories of the OLS, RR, WLS_G, and WLS_{RQ} estimators of β_0–β_3 in the HAR model for daily SPX raw RV, obtained using a rolling window. The end of the first 1000 day window used for parameter estimation is 6 April 2001, and the end of the last is 29 August 2013. As expected, there is substantial time variation in the rolling estimates. It is clear that although the estimates of the alternative estimation schemes often show a similar pattern, OLS estimates are much more variable than those of RR or WLS. This is most evident during the period 2006–2008 and then again towards the end of 2012. For these periods, associated with windows exhibiting both high and low volatility, RR and WLS estimates change more smoothly. This is due to RR and WLS being less sensitive to outlying observations than OLS. Whether these more stable HAR parameter estimates are of practical importance remains an empirical question to be answered in the out-of-sample forecasting exercise below.

The out-of-sample forecasting results are reported in Table 4 for SPX, and Table 5 for the sample of 26 individual stocks. In both tables, results are grouped by Benchmarks (HAR and HARQ), Estimators (RR and WLS), Transformations (log and sqr) and Combinations (the combinations of log, sqr, RR and WLS). Table 4 reports out-of-sample QLIKE, MSE and VaR loss ratios of the alternative approaches relative to the original HAR model, for 1-, 5-, 10- and 22-day forecast horizons. These ratios will be used to compare the performance of the alternative estimation, transformation, and combination schemes to the HAR and HARQ benchmarks (both estimated by OLS). To highlight the best performing approach, the lowest ratio in each row is shown in bold and approaches included in $\widehat{\mathcal{M}}^*_{90}$, a 90% MCS, are indicated by asterisks. As discussed earlier, for ease of comparison to the results of Bollerslev et al. (2016), a set of forecast comparisons with the IF (described in Section 2.6.3) applied to all forecasts are presented here. Insanity filters for volatility forecasting have also been used by Patton and Sheppard (2015) and Bollerslev et al. (2018a). Even so, the vast majority of studies examining the HAR model, and extensions thereof, do not employ an IF. Examples include Corsi and Renò (2012), Taylor (2017) and Cipollini et al. (2021). To examine the practical impact of employing an IF, results are reported for the case with the IF applied to all forecasts (IF on), and the case with no IF (IF off). In the latter case, the HARQ model is excluded due to its reliance on the IF (Bollerslev et al., 2016). The results for the individual stocks in Table 5 are structured in the same way with average loss ratios for the stocks reported, along with the number of times (in parentheses) each approach is included in $\widehat{\mathcal{M}}^*_{90}$.

3.3.1. The estimators

The goal of this section is to study the performance of alternative estimators for the HAR model. More specifically, to investigate how the predictive accuracy of the HAR depends on the choice of estimator. We start by comparing the loss ratios for SPX reported under Benchmarks and Estimators in columns 3–7 of Table 4. The first thing to note is that while the use of the IF does have an impact on the ratios to some degree, it has no substantial impact on the relative performance of the five approaches. At the 1-day horizon, the IF is not triggered for any approaches other than the HARQ, and hence has no impact. At the longer horizons, there are a small number of IF triggers. These occur during large, rapid increases in volatility during the onset of market turbulence where forecasts exceed values observed throughout the preceding estimation window. As the focus of this paper is on the practical application of the HAR model, a thorough investigation into how the HAR forecasts respond to rapid changes in volatility is beyond its scope. What is important here is that the choice of whether or not to use an IF has little overall impact on the relative performance of the alternative approaches. At the 1-day horizon, the RR-HAR and the two WLS schemes often offer substantial reductions in the loss measures compared to the HAR. The WLS_G-HAR has the lowest QLIKE. The HARQ has the lowest MSE, with the RR-HAR and WLS_G-HAR approaches not far behind, and the HAR has the lowest VaR. For the longer (5-, 10- and 22-day) horizons, a systematic pattern emerges. The RR-HAR and the two WLS schemes generally provide considerable reductions in QLIKE, MSE and VaR loss compared to the HAR. With the exception of the case with no IF and QLIKE loss, the RR-HAR consistently provides the lowest loss measures, outperforming both the HAR and HARQ. For the case with no IF, the WLS_{RQ}-HAR consistently has the lowest QLIKE. The corresponding results for the individual stocks, reported in columns 3–7 of Table 5, are remarkably similar. Except for the case with no

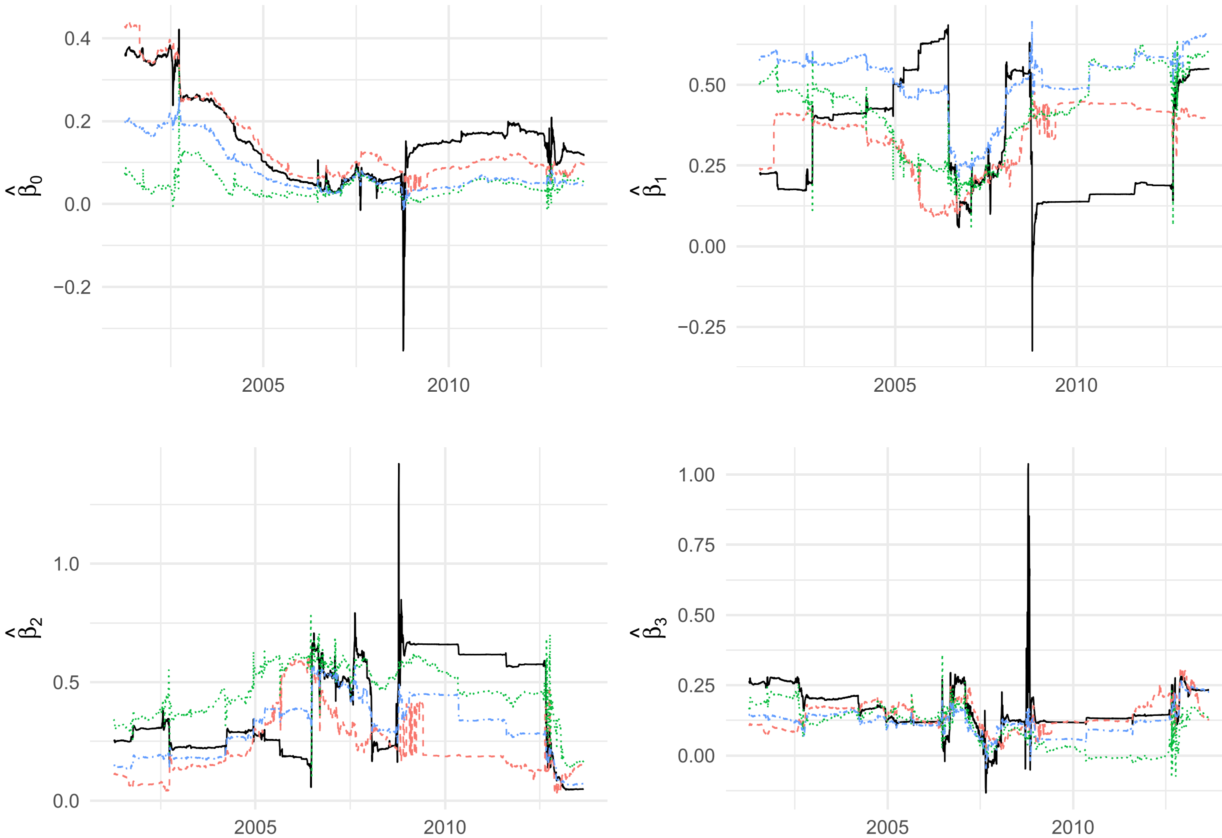

Fig. 2. Sample paths of the OLS (solid), RR (dashed), WLS_G (dotted), and WLS_{RQ} (dotdash) estimators of β_0–β_3 in the HAR volatility model for daily SPX raw RV, obtained using a 1000 day rolling window. Top panels: Trajectories of the estimators of β_0 and β_1. Bottom panels: Trajectories of the estimators of β_2 and β_3.

Table 4
Relative QLIKEs, MSEs, and VaRs for the HAR(Q) based out-of-sample volatility forecasts at 1-, 5-, 10- and 22-day horizons, obtained using alternative estimation, transformation, or combination schemes and a rolling window for SPX: QLIKE, MSE, and VaR ratios of the HAR to alternative approaches. The lowest ratio in each row is indicated in bold. IF on/off indicates if the "insanity filter" of Bollerslev et al. (2016) was applied or not. Asterisks indicate approaches included in $\widehat{\mathcal{M}}^*_{90}$.

		Benchmarks		Estimators			Transformations		Combinations					
		HAR	HARQ	RR-HAR	WLS_{RQ}-HAR	WLS_G-HAR	log-HAR	sqr-HAR	RR-log-HAR	RR-sqr-HAR	WLS_{RQ}-log-HAR	WLS_{RQ}-sqr-HAR	WLS_G-log-HAR	WLS_G-sqr-HAR
							1-day							
QLIKE	IF on	1.000	1.017	1.004	0.900*	**0.890***	0.898*	0.988	0.900*	1.003	0.898*	0.985	0.901*	0.973
	IF off	1.000	-	1.004	0.900*	**0.890***	0.898*	0.988	0.900*	1.003	0.898*	0.985	0.901*	0.973
MSE	IF on	1.000*	0.827*	0.873*	0.958*	0.878*	0.792*	0.848*	0.792*	0.844*	0.794*	0.832*	**0.784***	0.797*
	IF off	1.000*	-	0.873*	0.958*	0.878*	0.792*	0.848*	0.792*	0.844*	0.794*	0.832*	**0.784***	0.797*
VaR	IF on	**1.000***	1.010*	1.060	1.006*	1.003*	1.018	1.003*	1.025	1.033	1.016	1.004*	1.020	1.004*
	IF off	**1.000***	-	1.060	1.006*	1.003*	1.018	1.003*	1.025	1.033*	1.016	1.004*	1.020	1.004*
							5-day							
QLIKE	IF on	1.000	0.921	0.853	1.055	0.930	**0.795***	0.849	0.806*	0.830*	**0.795***	0.838*	0.820*	0.855
	IF off	1.000	-	0.941	**0.810***	0.972	0.871	0.872	0.892	0.921	0.865	0.854	0.995	0.997
MSE	IF on	1.000*	1.017*	0.864*	1.059*	0.981*	**0.843***	0.890*	0.852*	0.866*	**0.843***	0.872*	0.868*	0.885*
	IF off	1.000*	-	0.699*	0.802*	0.829*	0.667*	0.704*	0.685*	0.709*	**0.665***	**0.665***	0.758*	0.727*
VaR	IF on	1.000	0.985	**0.918***	0.977	0.963	0.932	0.970	0.923*	0.939	0.935	0.969	0.926*	0.956
	IF off	1.000	-	**0.908***	0.967	0.956	0.918	0.962	**0.908***	0.927	0.921	0.961	0.913*	0.947
							10-day							
QLIKE	IF on	1.000	0.931	0.735*	0.938*	0.959	0.741*	0.860*	0.675*	**0.654***	0.745*	0.870*	0.668*	0.786*
	IF off	1.000	-	0.956	**0.812***	0.981	0.894	0.859	0.932	0.955	0.887	0.842*	1.031	0.979
MSE	IF on	1.000*	0.999*	0.862*	1.013*	0.999*	0.886*	0.952*	0.837*	**0.827***	0.888*	0.958*	0.844*	0.905*
	IF off	1.000*	-	0.635*	0.723*	0.762*	0.600*	0.624*	0.635*	0.672*	0.598*	**0.585***	0.678*	0.617*
VaR	IF on	1.000	0.993	0.902*	0.971	0.959	0.908	0.958	**0.898***	0.915	0.910	0.960	0.902	0.944
	IF off	1.000	-	0.891	0.964	0.949	0.893	0.947	**0.879***	0.898	0.896	0.947	0.882	0.927
							22-day							
QLIKE	IF on	1.000	0.886*	0.799*	0.950*	0.981*	0.812*	0.883*	**0.753***	0.755*	0.810*	0.886*	0.796*	0.867*
	IF off	1.000	-	1.007	**0.829***	0.961	0.930	0.872	0.987	0.943	0.923	0.849*	1.023	0.939
MSE	IF on	1.000*	0.969*	0.861*	1.014*	1.015*	0.906*	0.964*	**0.839***	0.842*	0.905*	0.971*	0.880*	0.943*
	IF off	1.000*	-	0.749*	0.933*	0.889*	0.743*	0.741*	0.793*	0.741*	0.740*	**0.717***	0.785*	0.730*
VaR	IF on	1.000	0.991	0.887	0.959	0.952	0.889	0.941	**0.873***	0.903	0.892	0.941	0.881	0.927
	IF off	1.000	-	0.864	0.963	0.951	0.863	0.931	**0.839***	0.876	0.866	0.931	0.852	0.911

Table 5

Average relative QLIKEs, MSEs, and VaRs for the HAR(Q) based out-of-sample volatility forecasts at 1-, 5-, 10- and 22-day horizons, obtained using alternative estimation, transformation, or combination schemes and a rolling window for the 26 NYSE stocks: Average QLIKE, MSE, and VaR ratios of the HAR to alternative approaches. The lowest ratio in each row is indicated in bold. IF on/off indicates if the "insanity filter" of Bollerslev et al. (2016) was applied or not. Numbers in parentheses indicate how many times each approach is included in $\widehat{\mathcal{M}}^*_{90}$.

		Benchmarks		Estimators			Transformations		Combinations					
		HAR	HARQ	RR-HAR	WLS_{RQ}-HAR	WLS_G-HAR	log-HAR	sqr-HAR	RR-log-HAR	RR-sqr-HAR	WLS_{RQ}-log-HAR	WLS_{RQ}-sqr-HAR	WLS_G-log-HAR	WLS_G-sqr-HAR
							1-day							
QLIKE	IF on	1.000$^{(1)}$	1.032$^{(6)}$	1.056$^{(0)}$	0.939$^{(23)}$	0.944$^{(19)}$	0.934$^{(26)}$	0.964$^{(6)}$	0.943$^{(17)}$	0.992$^{(0)}$	**0.930**$^{(26)}$	0.963$^{(7)}$	0.934$^{(26)}$	0.961$^{(7)}$
	IF off	1.000$^{(1)}$	-	1.064$^{(0)}$	0.946$^{(23)}$	0.952$^{(19)}$	0.941$^{(26)}$	0.971$^{(5)}$	0.950$^{(17)}$	0.999$^{(0)}$	**0.937**$^{(26)}$	0.971$^{(6)}$	0.941$^{(26)}$	0.968$^{(7)}$
MSE	IF on	1.000$^{(25)}$	0.958$^{(26)}$	0.953$^{(25)}$	0.983$^{(26)}$	0.923$^{(25)}$	0.867$^{(26)}$	0.893$^{(26)}$	0.878$^{(26)}$	0.925$^{(26)}$	**0.864**$^{(26)}$	0.881$^{(26)}$	0.869$^{(26)}$	0.869$^{(26)}$
	IF off	1.000$^{(25)}$	-	0.924$^{(25)}$	0.951$^{(26)}$	0.894$^{(25)}$	0.840$^{(26)}$	0.865$^{(26)}$	0.852$^{(26)}$	0.897$^{(26)}$	**0.838**$^{(26)}$	0.854$^{(26)}$	0.842$^{(26)}$	0.843$^{(26)}$
VaR	IF on	1.000$^{(18)}$	1.001$^{(23)}$	0.994$^{(26)}$	0.995$^{(25)}$	0.993$^{(25)}$	0.993$^{(25)}$	0.997$^{(22)}$	0.993$^{(25)}$	0.996$^{(25)}$	0.993$^{(25)}$	0.998$^{(20)}$	**0.992**$^{(25)}$	0.996$^{(22)}$
	IF off	1.000$^{(19)}$	-	0.994$^{(26)}$	0.995$^{(25)}$	0.993$^{(25)}$	0.993$^{(25)}$	0.997$^{(23)}$	0.993$^{(25)}$	0.996$^{(25)}$	0.993$^{(25)}$	0.998$^{(21)}$	**0.992**$^{(25)}$	0.996$^{(22)}$
							5-day							
QLIKE	IF on	1.000$^{(2)}$	0.936$^{(9)}$	0.909$^{(6)}$	0.941$^{(11)}$	0.969$^{(1)}$	0.845$^{(14)}$	0.884$^{(17)}$	0.842$^{(18)}$	0.853$^{(17)}$	**0.840**$^{(25)}$	0.867$^{(20)}$	0.884$^{(12)}$	0.922$^{(2)}$
	IF off	1.000$^{(0)}$	-	1.032$^{(0)}$	**0.913**$^{(26)}$	1.061$^{(0)}$	0.967$^{(1)}$	0.928$^{(23)}$	0.998$^{(0)}$	0.987$^{(4)}$	0.953$^{(23)}$	0.917$^{(25)}$	1.090$^{(0)}$	1.056$^{(0)}$
MSE	IF on	1.000$^{(26)}$	0.962$^{(26)}$	0.839$^{(26)}$	0.950$^{(26)}$	0.907$^{(26)}$	0.803$^{(26)}$	0.868$^{(26)}$	**0.802**$^{(26)}$	0.819$^{(26)}$	0.804$^{(26)}$	0.847$^{(26)}$	0.823$^{(26)}$	0.862$^{(26)}$
	IF off	1.000$^{(26)}$	-	0.826$^{(25)}$	0.878$^{(26)}$	0.884$^{(24)}$	0.800$^{(25)}$	0.793$^{(26)}$	0.840$^{(25)}$	0.827$^{(25)}$	0.794$^{(26)}$	**0.768**$^{(26)}$	0.904$^{(25)}$	0.851$^{(25)}$
VaR	IF on	1.000$^{(0)}$	0.988$^{(0)}$	0.923$^{(19)}$	0.988$^{(0)}$	0.961$^{(0)}$	0.932$^{(0)}$	0.970$^{(0)}$	**0.920**$^{(25)}$	0.937$^{(0)}$	0.937$^{(0)}$	0.976$^{(0)}$	0.923$^{(14)}$	0.954$^{(0)}$
	IF off	1.000$^{(0)}$	-	0.916$^{(19)}$	0.988$^{(0)}$	0.957$^{(0)}$	0.923$^{(0)}$	0.965$^{(0)}$	**0.911**$^{(25)}$	0.929$^{(1)}$	0.928$^{(0)}$	0.970$^{(0)}$	0.913$^{(19)}$	0.947$^{(0)}$
							10-day							
QLIKE	IF on	1.000$^{(4)}$	0.934$^{(10)}$	0.876$^{(14)}$	0.960$^{(13)}$	0.970$^{(5)}$	0.862$^{(8)}$	0.898$^{(17)}$	0.821$^{(21)}$	**0.818**$^{(23)}$	0.855$^{(24)}$	0.886$^{(19)}$	0.863$^{(11)}$	0.900$^{(12)}$
	IF off	1.000$^{(2)}$	-	1.033$^{(3)}$	**0.895**$^{(26)}$	1.045$^{(0)}$	0.997$^{(2)}$	0.923$^{(19)}$	1.049$^{(1)}$	1.013$^{(5)}$	0.983$^{(9)}$	0.910$^{(26)}$	1.126$^{(0)}$	1.043$^{(0)}$
MSE	IF on	1.000$^{(26)}$	0.981$^{(26)}$	0.874$^{(26)}$	1.013$^{(26)}$	0.977$^{(26)}$	0.870$^{(26)}$	0.940$^{(26)}$	**0.827**$^{(26)}$	0.853$^{(26)}$	0.872$^{(26)}$	0.930$^{(26)}$	0.850$^{(26)}$	0.913$^{(26)}$
	IF off	1.000$^{(26)}$	-	0.775$^{(26)}$	0.827$^{(26)}$	0.836$^{(26)}$	0.757$^{(26)}$	0.739$^{(26)}$	0.816$^{(26)}$	0.810$^{(26)}$	0.753$^{(26)}$	**0.713**$^{(26)}$	0.852$^{(26)}$	0.775$^{(26)}$
VaR	IF on	1.000$^{(0)}$	0.994$^{(0)}$	0.919$^{(14)}$	0.987$^{(0)}$	0.963$^{(0)}$	0.923$^{(0)}$	0.962$^{(0)}$	**0.911**$^{(26)}$	0.928$^{(3)}$	0.927$^{(0)}$	0.967$^{(0)}$	0.914$^{(15)}$	0.948$^{(0)}$
	IF off	1.000$^{(0)}$	-	0.904$^{(7)}$	0.986$^{(0)}$	0.956$^{(0)}$	0.906$^{(0)}$	0.954$^{(0)}$	**0.890**$^{(26)}$	0.910$^{(3)}$	0.910$^{(0)}$	0.958$^{(0)}$	0.893$^{(10)}$	0.934$^{(0)}$
							22-day							
QLIKE	IF on	1.000$^{(15)}$	0.957$^{(23)}$	0.903$^{(25)}$	0.979$^{(23)}$	0.968$^{(20)}$	0.922$^{(18)}$	0.922$^{(24)}$	0.894$^{(23)}$	**0.867**$^{(25)}$	0.915$^{(24)}$	0.919$^{(25)}$	0.930$^{(19)}$	0.914$^{(24)}$
	IF off	1.000$^{(9)}$	-	1.025$^{(16)}$	**0.900**$^{(26)}$	1.015$^{(4)}$	1.060$^{(5)}$	0.954$^{(20)}$	1.128$^{(3)}$	1.039$^{(13)}$	1.049$^{(8)}$	0.935$^{(26)}$	1.156$^{(1)}$	1.034$^{(7)}$
MSE	IF on	1.000$^{(25)}$	0.985$^{(26)}$	0.897$^{(26)}$	1.020$^{(25)}$	0.985$^{(26)}$	0.919$^{(26)}$	0.959$^{(26)}$	**0.874**$^{(26)}$	0.879$^{(26)}$	0.918$^{(26)}$	0.959$^{(26)}$	0.902$^{(26)}$	0.929$^{(26)}$
	IF off	1.000$^{(25)}$	-	0.816$^{(26)}$	0.916$^{(26)}$	0.862$^{(25)}$	0.850$^{(25)}$	0.825$^{(25)}$	0.910$^{(25)}$	0.867$^{(26)}$	0.849$^{(25)}$	**0.799**$^{(25)}$	0.906$^{(25)}$	0.831$^{(25)}$
VaR	IF on	1.000$^{(0)}$	0.995$^{(0)}$	0.911$^{(7)}$	0.982$^{(0)}$	0.964$^{(0)}$	0.911$^{(0)}$	0.952$^{(0)}$	**0.896**$^{(26)}$	0.917$^{(2)}$	0.913$^{(0)}$	0.956$^{(0)}$	0.903$^{(4)}$	0.939$^{(0)}$
	IF off	1.000$^{(0)}$	-	0.894$^{(4)}$	0.986$^{(0)}$	0.959$^{(0)}$	0.888$^{(0)}$	0.944$^{(0)}$	**0.867**$^{(26)}$	0.896$^{(2)}$	0.891$^{(0)}$	0.947$^{(0)}$	0.877$^{(1)}$	0.924$^{(0)}$

IF and QLIKE loss, where the WLS_{RQ}-HAR has the lowest average QLIKE at all forecast horizons, the RR-HAR consistently provides the lowest average QLIKE, MSE and VaR at longer horizons. In sum, the results here suggest that the choice of the estimation scheme under which the HAR parameter estimates are obtained is important for volatility forecasting. This is perhaps not surprising given that the HAR parameter estimates obtained under the rolling RR and WLS schemes are more stable over time compared to the corresponding OLS estimates (cf. Fig. 2). This is often associated with more accurate forecasts.

3.3.2. The transformations

The aim of this section is to investigate whether transforming RV leads to an improvement in predictive accuracy over the benchmark models. As before, log-HAR denotes the standard HAR model fitted to the natural logarithm of RV, and sqr-HAR the same model fitted to the square root transformation of RV. Both of these (nonlinear) models for RV are estimated using OLS. We first compare the loss ratios for SPX reported under Benchmarks and Transformations in columns 3–4 and 8–9 of Table 4. At the 1-day horizon, the two transformation schemes often offer substantial reductions in the loss measures compared to the HAR. The log-HAR has the lowest QLIKE and MSE, and the HAR has the lowest VaR. At the longer horizons, the log-HAR and sqr-HAR consistently outperform the benchmarks under all three loss functions. Overall, the logarithmic transformation does particularly well. The sqr-HAR occasionally does better for the case with no IF. The corresponding results for the individual stocks, reported in columns 3–4 and 8–9 of Table 5, are much the same. At all four forecast horizons, the log-HAR and sqr-HAR outperform the HAR and HARQ in terms of average QLIKE, MSE and VaR loss. While the log-HAR still does particularly well overall, the sqr-HAR often does better for the case with no IF. To sum up, it is clear that transformations of RV also can have a significant impact on predictive accuracy.

3.3.3. Combinations

This section considers the possible benefits in predictive accuracy of combinations of different estimation (RR or WLS) and transformation (log or sqr) schemes. We start by comparing the loss ratios for SPX reported under Transformations and Combinations in columns 8–9 and 10–15 of Table 4. Here, for example, WLS_G-log-HAR denotes the standard HAR model fitted to the natural logarithm of RV using the WLS_G scheme. At the 1- and 5-day horizons, the six combination schemes offer only minor reductions, if any, in the loss measures compared to the two transformation schemes. At the 10- and 22-day horizons, the benefits of the combination schemes are more clear. The RR-log-HAR does particularly well overall. However, the WLS_{RQ}-sqr-HAR often does better for the case with no IF. While WLS may be useful for square root transformed RV, the two combination schemes using WLS and logarithmic RV generally offer minor reductions, if any, in the loss measures compared to the log-HAR. This indicates that the logarithmic transformation deals with the heteroskedasticity better than the square root transformation. The corresponding results for the individual stocks, reported in columns 8–9 and 10–15 of Table 5, are very similar. At the 1- and 5-day horizons, the combination schemes on average provide only minor to moderate reductions, if any, in the loss measures compared to the log-HAR and sqr-HAR. At the 10- and 22-day horizons, however, the benefits of the combination schemes are more apparent. Overall, the RR-log-HAR does particularly well. However, the WLS_{RQ}-sqr-HAR frequently does better for the case with no IF. In sum, the results here suggest that there can be notable benefits from estimating HAR models for transformed RV using estimation schemes other than OLS.

3.3.4. Estimators, transformations, or combinations?

Finally a comparison across all of the approaches is presented to determine whether the choice of estimator, transformation, or combination thereof, provide the greatest benefit for forecasting volatility with the HAR model. Here the best performing approaches will be discussed in terms of the loss measures and the MCS. Once again we start with the results for SPX in Table 4, where the lowest loss ratio in each row is indicated in bold and asterisks indicate approaches included in a 90% MCS. Overall, the RR-log-HAR stands out the most. It is typically included in the MCS, and most often has the lowest loss ratios. The WLS_{RQ}-HAR and its combinations, WLS_{RQ}-log-HAR and WLS_{RQ}-sqr-HAR, also do well overall. Notably, these approaches all work best at longer forecast horizons. With the exception of MSE loss, the benchmark HAR and HARQ models are typically not included in the MCS. The corresponding results for the 26 NYSE stocks in Table 5 are strikingly similar. Once again the RR-log-HAR does best overall, and the WLS_{RQ}-HAR and its combinations, WLS_{RQ}-log-HAR and WLS_{RQ}-sqr-HAR, also stand out. To conclude, moving away from OLS in terms of estimation, or transforming RV, generally offers clear improvements in predictive accuracy. However, the results here show that further improvements (although typically smaller in magnitude than the initial gains provided by RR, WLS, log or sqr) can be obtained by estimating HAR models for transformed RV using estimators different from OLS. Overall, these combination schemes produce the best performing forecasts, with or without an insanity filter.

3.3.5. Over/under-prediction

The aim of this section is to further investigate the interaction between forecasts from the alternative approaches and the loss functions in Section 2.7.1 used for forecast comparison. These loss functions deal with over- and under-prediction differently. While it is difficult to make general claims about preferences for loss functions, to some practitioners, symmetric loss functions, such as MSE, may be less attractive than asymmetric ones. Under asymmetric loss functions, such as QLIKE, negative forecast errors (over-predictions) are penalized differently from positive forecast errors (under-predictions).

In view of this, we first examine over- and under-prediction. Table 6 reports out-of-sample proportions of over-prediction (POPs), mean over-predictions (MOPs) and mean under-predictions (MUPs) for the alternative approaches and 1-, 5-, 10- and 22-day forecast horizons for SPX. The POPs indicate that while all approaches are more likely to over-predict than to under-predict, some are less likely to over-predict than others. Replacing OLS with RR or WLS typically leads to approaches that are less likely to over-predict. The RR-HAR, for example, has lower POP than the HAR, HARQ and sqr-HAR for all forecast horizons – with or without an insanity filter. Overall, the POPs increase with the forecast horizon, with the RR-log-HAR exhibiting the lowest POPs. For the 1-day forecast horizon, the POPs for all approaches remain the same after the insanity filter has been applied to all forecasts. For the 5-, 10- and 22-day forecast horizons, the IF typically reduces the POPs, but not for the schemes based on logarithmic RV.[7] The MOPs suggest that the average negative forecast error also varies across the alternative approaches. Replacing OLS with RR or WLS typically leads to approaches with lower average absolute negative forecast errors. The RR-HAR, as an example, has lower average absolute over-prediction than the HAR, HARQ, log-HAR and sqr-HAR for all forecast horizons – with or without an insanity filter. The IF reduces the average absolute over-predictions for the 5-, 10- and 22-day forecast horizons in most cases. The MUPs indicate that the average positive forecast error varies across the alternative approaches as well. Replacing OLS with RR or WLS typically leads to approaches with lower average positive forecast errors. Overall, schemes based on logarithmic RV, such as RR-log-HAR, have the lowest MUPs. In all but one case, the IF employed by Bollerslev et al. (2016) inflates the MUPs for the 5-, 10- and 22-day forecast horizons.

To better understand the benefits of replacing OLS with a more suitable estimation method, we next focus on some selected approaches. Fig. 3 shows grouped (IF on/off) box plots for six of the alternative approaches, visualizing five summary statistics (the median, two hinges and two whiskers) for SPX 5-day horizon forecast errors (target RV – forecast, as in the definition of MSE) and forecast ratios (target RV/forecast, as in the definition of QLIKE), with respective MSE and QLIKE loss functions superimposed.[8] The box plots provide additional clues to help explain the good out-of-sample performance of the RR-HAR, WLS_{RQ}-HAR and log-HAR. For these approaches, the bulk of the forecast errors are shifted closer to zero (top panel, MSE) and forecast ratios closer to one (bottom panel, QLIKE). That is, into regions of lower loss. Under QLIKE, the RR-HAR, WLS_{RQ}-HAR and log-HAR have less (more) forecast ratios in the more (less) heavily penalized region of the domain of the loss function. Results for the 1-, 10- and 22-day forecast horizons are broadly similar to those reported here and are available in the Online Appendix. While the (OLS-)HAR excessively over-predicts, in particular for the longer horizons, the alternative estimation schemes (RR, WLS) partially reduce this asymmetry.

In addition to MSE and QLIKE, a VaR-based loss function was used for forecast comparison. However, it is not possible to discuss VaR in the same manner as MSE and QLIKE since the former loss function does not involve target RV. See Eqs. (13) and (14).

3.3.6. Robustness checks

This section presents three robustness checks. First, for the SPX dataset, forecast performance will be compared over high and low volatility periods. Second, the performance of a few selected approaches will be compared to some extended HAR models previously suggested in the literature. As the underlying intraday 5-minute returns are not available for this specific SPX dataset, this part of the analysis will be based on the 26 individual stocks. Third, indirect forecasts from some selected approaches for the SPX dataset will be evaluated and compared to the corresponding direct forecasts.

First, two subsamples of the SPX series corresponding to periods of high and low volatility respectively, were selected to examine the robustness of the results with respect to changing market conditions. This will reveal if the performance of the alternative estimation, transformation and combination schemes reported in sections 3.3.1–3.3.4 are influenced by changes in market conditions. The two subsamples split the full out-of-sample period

[7] In most cases, the IF described in Section 2.6.3 changes a small number of very large forecasts: any forecast greater than the maximum of the dependent variable observed in the estimation period is replaced by the sample average over that period. Thus POP is generally reduced. An increase in POP after application of the IF is also possible as the IF in some cases changes a small number of very small forecasts: any forecast less than the minimum of the dependent variable observed in the estimation period is replaced by the sample average over that period. This can increase POP.

[8] In these Tukey style box and whiskers plots, lower and upper hinges correspond to first and third quartiles (Q_1 and Q_3). Upper whiskers extend from each upper hinge to the largest value no further than $1.5 \times IQR$ from the hinge, where $IQR = Q_3 - Q_1$ is the interquartile range. Similarly, lower whiskers extend from each lower hinge to the smallest value at most $1.5 \times IQR$ from the hinge. For ease of exposition, values beyond the end of the whiskers (outliers) are not shown.

Table 6
POPs (proportions of over-prediction), MOPs (mean over-predictions), and MUPs (mean under-predictions) for the HAR(Q) based out-of-sample volatility forecasts at 1-, 5-, 10- and 22-day horizons, obtained using alternative estimation, transformation, or combination schemes and a rolling window for SPX. IF on/off indicates if the "insanity filter" of Bollerslev et al. (2016) was applied or not.

		Benchmarks		Estimators			Transformations		Combinations					
		HAR	HARQ	RR-HAR	WLS_{RQ}-HAR	WLS_G-HAR	log-HAR	sqr-HAR	RR-log-HAR	RR-sqr-HAR	WLS_{RQ}-log-HAR	WLS_{RQ}-sqr-HAR	WLS_G-log-HAR	WLS_G-sqr-HAR
							1-day							
POP	IF on	0.693	0.613	0.542	0.649	0.631	0.625	0.698	0.612	0.655	0.631	0.707	0.623	0.698
	IF off	0.693	-	0.542	0.649	0.631	0.625	0.698	0.612	0.655	0.631	0.707	0.623	0.698
MOP	IF on	−0.411	−0.427	−0.237	−0.401	−0.421	−0.344	−0.372	−0.319	−0.280	−0.354	−0.370	−0.337	−0.365
	IF off	−0.411	-	−0.237	−0.401	−0.421	−0.344	−0.372	−0.319	−0.280	−0.354	−0.370	−0.337	−0.365
MUP	IF on	0.726	0.573	0.689	0.619	0.574	0.635	0.732	0.650	0.796	0.631	0.737	0.639	0.723
	IF off	0.726	-	0.689	0.619	0.574	0.635	0.732	0.650	0.796	0.631	0.737	0.639	0.723
							5-day							
POP	IF on	0.734	0.685	0.565	0.666	0.653	0.555	0.678	0.537	0.618	0.562	0.675	0.557	0.654
	IF off	0.740	-	0.566	0.672	0.657	0.554	0.681	0.535	0.620	0.561	0.678	0.556	0.656
MOP	IF on	−0.351	−0.364	−0.213	−0.336	−0.318	−0.265	−0.297	−0.246	−0.215	−0.272	−0.298	−0.252	−0.277
	IF off	−0.409	-	−0.218	−0.379	−0.345	−0.264	−0.317	−0.244	−0.217	−0.272	−0.313	−0.249	−0.282
MUP	IF on	0.963	0.785	0.737	0.783	0.815	0.644	0.816	0.648	0.809	0.645	0.797	0.701	0.845
	IF off	0.766	-	0.671	0.598	0.695	0.579	0.693	0.588	0.740	0.577	0.676	0.650	0.753
							10-day							
POP	IF on	0.762	0.730	0.586	0.701	0.684	0.564	0.683	0.540	0.616	0.572	0.681	0.557	0.663
	IF off	0.771	-	0.587	0.709	0.692	0.559	0.688	0.534	0.617	0.568	0.685	0.553	0.665
MOP	IF on	−0.340	−0.348	−0.209	−0.321	−0.296	−0.252	−0.285	−0.230	−0.201	−0.257	−0.289	−0.238	−0.265
	IF off	−0.444	-	−0.212	−0.389	−0.360	−0.255	−0.323	−0.221	−0.201	−0.260	−0.317	−0.233	−0.277
MUP	IF on	1.323	1.122	0.859	1.034	1.075	0.769	1.006	0.747	0.915	0.780	0.984	0.783	0.999
	IF off	0.922	-	0.757	0.697	0.802	0.650	0.757	0.669	0.841	0.654	0.744	0.716	0.816
							22-day							
POP	IF on	0.771	0.755	0.615	0.737	0.739	0.594	0.714	0.564	0.653	0.599	0.709	0.587	0.700
	IF off	0.785	-	0.619	0.751	0.752	0.588	0.721	0.556	0.654	0.593	0.716	0.583	0.704
MOP	IF on	−0.401	−0.392	−0.229	−0.351	−0.329	−0.265	−0.308	−0.228	−0.235	−0.269	−0.314	−0.253	−0.288
	IF off	−0.501	-	−0.232	−0.451	−0.408	−0.266	−0.351	−0.218	−0.235	−0.269	−0.352	−0.248	−0.309
MUP	IF on	1.689	1.547	1.072	1.465	1.558	1.006	1.384	0.952	1.166	1.014	1.348	0.996	1.334
	IF off	1.170	-	0.935	0.977	1.107	0.820	1.018	0.850	1.024	0.823	0.984	0.861	1.040

nearly exactly in half. The low volatility subsample starts at the beginning of the out-of-sample period and spans 9 April 2001 to 19 April 2007. The high volatility subsample spans 20 April 2007 to 30 August 2013, a time period which contains the global financial crisis, a period of market turbulence and record high levels of volatility. Table 7 presents forecasting results for SPX over the high and low volatility subsamples. To conserve space, only results for the case with the IF applied to all forecasts are reported. Overall, the RR-log-HAR performs the best. It is generally included in the MCS, and most often has the lowest loss ratios. The log-HAR and WLS_{RQ}-HAR also do well overall. As expected, the RR-HAR stands out over the high volatility subsample. The benchmark HAR and HARQ models are typically not included in the MCS. The alternative estimation, transformation and combination schemes are on the whole superior to the benchmarks over both subsamples. Overall, the results here suggest that these schemes are robust to changes in market conditions in terms of the level of volatility.

Next, a selection of the approaches considered in sections 3.3.1–3.3.2 together with the HAR and HARQ are compared to some extended HAR models suggested in the literature. The first extension is a leverage HAR (LHAR) model considered in Corsi and Renò (2012). The LHAR model can be expressed as

$$\log RV_t = \beta_0 + \beta_1 \log RV^d_{t-1} + \beta_2 \log RV^w_{t-1} + \beta_3 \log RV^m_{t-1} + \beta_4 r^{d-}_{t-1} + \beta_5 r^{w-}_{t-1} + \beta_6 r^{m-}_{t-1} + u_t,$$

where $\log RV_t$ is the logarithmic RV of day t, and $\log RV^d_{t-1} = \log RV_{t-1}$, $\log RV^w_{t-1} = \frac{1}{5}\sum_{i=1}^{5} \log RV_{t-i}$, $\log RV^m_{t-1} = \frac{1}{22}\sum_{i=1}^{22} \log RV_{t-i}$ denote the daily, weekly and monthly lagged logarithmic RV, respectively, r_t is the daily return, and $r^{d-}_{t-1} = \min(r^d_{t-1}, 0)$, $r^{w-}_{t-1} = \min(r^w_{t-1}, 0)$, $r^{m-}_{t-1} = \min(r^m_{t-1}, 0)$, with $r^d_{t-1} = r_{t-1}$, $r^w_{t-1} = \frac{1}{5}\sum_{i=1}^{5} r_{t-i}$, $r^m_{t-1} = \frac{1}{22}\sum_{i=1}^{22} r_{t-i}$. The second extension considered is the HAR-RSV model based on realized semi-variances (RSVs) of Chen and Ghysels (2011). Patton and Sheppard (2015) also use a similar structure. The RSV estimators

$$RV^+_t = \sum_{i=1}^{M} r^2_{t,i} \mathbf{1}_{r_{t,i}>0}, \quad \text{and} \quad RV^-_t = \sum_{i=1}^{M} r^2_{t,i} \mathbf{1}_{r_{t,i}\leq 0},$$

split daily RV into semi-variances. The HAR-RSV model can be written as

$$RV_t = \beta_0 + \beta_1 RV^{d+}_{t-1} + \beta_2 RV^{w+}_{t-1} + \beta_3 RV^{m+}_{t-1} + \beta_4 RV^{d-}_{t-1} + \beta_5 RV^{w-}_{t-1} + \beta_6 RV^{m-}_{t-1} + u_t,$$

where $RV^{d+}_{t-1} = RV^+_{t-1}$, $RV^{w+}_{t-1} = \frac{1}{5}\sum_{i=1}^{5} RV^+_{t-i}$, $RV^{m+}_{t-1} = \frac{1}{22}\sum_{i=1}^{22} RV^+_{t-i}$, and RV^{d-}_{t-1}, RV^{w-}_{t-1}, RV^{m-}_{t-1} are defined similarly. The third, and final, extension considered is the HAR-CJ model of Andersen et al. (2007), which harnesses the continuous component (C) and jump component (J) of RV. With daily, weekly and monthly lagged continuous and jump components, the HAR-CJ model can be expressed as

$$RV_t = \beta_0 + \beta_1 C^d_{t-1} + \beta_2 C^w_{t-1} + \beta_3 C^m_{t-1} + \beta_4 J^d_{t-1} + \beta_5 J^w_{t-1} + \beta_6 J^m_{t-1} + u_t.$$

Table 8 presents forecasting results for these extended HAR models along with results for the RR-HAR, WLS_{RQ}-HAR and log-HAR, and the two benchmark models. Here, the analysis is based on the sample of 26 individual stocks as the underlying intraday returns needed to compute the RSVs and the C and J components of RV are not publicly available for the SPX dataset. As expected, forecasts from the state of the art HARQ model, on average, generally outperform those from the HAR-RSV and HAR-CJ models in terms of QLIKE, MSE and VaR loss. In fact, at best, these two HAR extensions provide only minor to moderate reductions in the loss measures on average compared to the standard HAR. The LHAR based on logarithmic (instead of raw) RV, on the other hand, performs comparatively well. Overall, the log-HAR and RR-HAR con-

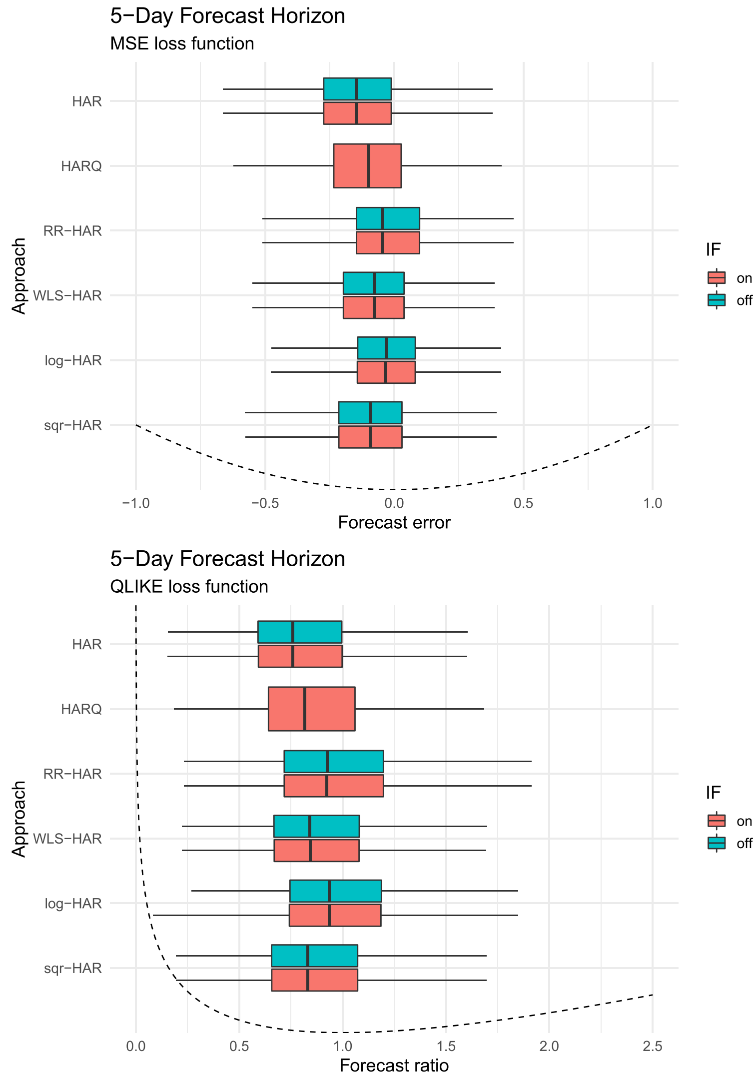

Fig. 3. Grouped box plots for six alternative approaches, with WLS-HAR indicating the WLS_{RQ}-HAR. For each group, top box plots represent IF off and bottom IF on (HARQ requires IF on). Top panel: SPX rolling window 5-day horizon forecast errors, with superimposed MSE loss function (dashed). Bottom panel: SPX rolling window 5-day horizon forecast ratios, with superimposed QLIKE loss function (dashed).

tinue to stand out as the best performers. They are both typically included in the MCS, and mostly have the lowest average loss ratios. WLS_{RQ}-HAR also does comparatively well, in general outperforming the HAR extensions based on raw RV. The focus of this paper has been on the benefits of various schemes for the standard HAR model. These last results show that forecasts generated by these schemes can outperform those from commonly used extensions of the HAR model, without resorting to using intraday returns. This observation is of practical interest as RV often is publicly available but the underlying intraday returns needed for computing the regressors of some of the HAR extensions are not.

For longer forecast horizons, an alternative to the direct method described in Section 2.6.1 is the indirect approach. See, for example, Chevillon and Hendry (2005) and Marcellino et al. (2006). In our setting, the indirect (iterative) approach first fits a daily RV model and then iterates over its daily (h-day-ahead) forecasts to obtain weekly, biweekly, or monthly predictions.[9] The aim of this part is to investigate if the indirect approach, similar to the direct approach, benefits from the alternative combination schemes for the HAR model and, in addition, if the indirect approach provides more accurate forecasts than the direct approach.

We first examine if the benefits of the alternative combination schemes are still observed under an indirect forecasting approach. Table 9 reports loss ratios for the 5-, 10- and 22-day forecast horizons (indirect and direct forecasts are identical at the 1-day horizon) and the RR, WLS_{RQ}, and WLS_{G} estimation schemes applied to the log-HAR model. Iterative bias-corrected h-day-ahead forecasts of RV for the log-HAR model were computed using Proposition 1 in Buccheri and Corsi (2019).[10] These daily forecasts were then properly aggregated to obtain the longer horizon predictions. To make the loss ratios directly comparable to those reported in 4, Table 9 reports QLIKE, MSE and VaR ratios of the (direct forecast)

[9] Direct forecasts are easy to compute and more robust to model misspecification compared to indirect forecasts. The indirect approach is also much more challenging for the HARQ and other extensions of the HAR model (such as the LHAR) as these models would require the dynamics of their exogenous variables to be specified. This often makes direct volatility forecasts preferable for longer forecast horizons.

[10] It appears that closed form expressions for iterative bias-corrected multistep-ahead forecasts under some other transformation schemes (such as the sqr-HAR) are not currently available. Deriving and evaluating such expressions may be an interesting avenue for future research.

Table 7
Relative QLIKEs, MSEs, and VaRs for the HAR(Q) based out-of-sample volatility forecasts at 1-, 5-, 10- and 22-day horizons, obtained using alternative estimation, transformation, or combination schemes, the "insanity filter" of Bollerslev et al. (2016), and a rolling window for SPX: QLIKE, MSE, and VaR ratios of the HAR to alternative approaches. The lowest ratio in each row is indicated in bold. High/Low indicates if the subsample used for estimation was a high/low volatility period. Asterisks indicate approaches included in $\widehat{\mathcal{M}}^*_{90}$.

		Benchmarks		Estimators			Transformations		Combinations					
		HAR	HARQ	RR-HAR	WLS_{RQ}-HAR	WLS_G-HAR	log-HAR	sqr-HAR	RR-log-HAR	RR-sqr-HAR	WLS_{RQ}-log-HAR	WLS_{RQ}-sqr-HAR	WLS_G-log-HAR	WLS_G-sqr-HAR
							1-day							
QLIKE	High	1.000	1.057	1.015	**0.878***	0.879*	0.890*	0.995	0.897*	1.023	0.888*	0.989	0.895*	0.985
	Low	1.000	0.953	0.986	0.935	0.908*	0.909*	0.977	**0.905***	0.972	0.914	0.979	0.909*	0.954
MSE	High	1.000*	0.821*	0.861*	0.963*	0.874*	0.784*	0.842*	0.784*	0.832*	0.787*	0.828*	**0.776***	0.790*
	Low	1.000	0.906*	1.042	**0.888***	0.940*	0.905*	0.938	0.920*	1.014	0.892*	0.896*	0.890*	0.901*
VaR	High	1.000*	1.012*	1.081	1.005*	**0.998***	1.025*	1.007*	1.034	1.052*	1.022*	1.009*	1.028	1.010*
	Low	1.000*	1.007	1.025	1.007	1.011	1.007	0.996*	1.010	1.003*	1.006	0.996*	1.007	**0.995***
							5-day							
QLIKE	High	1.000	0.906*	0.851*	1.122*	0.930*	**0.802***	0.858*	0.816*	0.836*	0.803*	0.852*	0.812*	0.846*
	Low	1.000	0.979	0.858	0.799*	0.932	**0.767***	0.812	0.771*	0.805*	**0.767***	0.784*	0.849	0.891
MSE	High	1.000*	1.022*	0.856*	1.072*	0.975*	**0.840***	0.888*	0.849*	0.860*	0.841*	0.872*	0.860*	0.875*
	Low	1.000	0.941	1.001	**0.850***	1.076	0.895*	0.914*	0.910*	0.955*	0.883*	0.869*	1.002	1.038
VaR	High	1.000	0.985	**0.903***	0.985	0.961	0.935	0.975	0.924	0.936	0.938	0.975	0.928	0.957
	Low	1.000	0.985	0.938	0.968	0.966	0.928	0.963	**0.920***	0.942	0.930	0.962	0.923*	0.955
							10-day							
QLIKE	High	1.000	0.922*	0.737*	0.970*	0.971*	0.747*	0.878*	0.667*	**0.643***	0.753*	0.896*	0.653*	0.782*
	Low	1.000	0.985	0.721*	0.747*	0.888	0.701*	0.752	0.724*	0.725*	**0.699***	0.716*	0.757*	0.806
MSE	High	1.000*	1.001*	0.859*	1.022*	0.996*	0.885*	0.954*	0.832*	**0.822***	0.888*	0.963*	0.836*	0.899*
	Low	1.000	0.955	0.911*	**0.845***	1.057	0.903*	0.901*	0.925*	0.917*	0.892*	0.856*	0.995*	1.005
VaR	High	1.000	0.997	**0.883***	0.977	0.958	0.908	0.966	0.895*	0.905	0.910	0.971	0.901	0.948
	Low	1.000	0.989	0.926	0.963	0.961	0.908	0.948	**0.902***	0.927	0.910	0.946	**0.902***	0.938
							22-day							
QLIKE	High	1.000*	0.882*	0.818*	0.991*	1.000*	0.840*	0.914*	**0.771***	0.772*	0.837*	0.926*	0.820*	0.892*
	Low	1.000	0.913	0.684*	0.700*	0.867	**0.643***	0.696	0.645*	0.648*	0.646*	0.649*	0.651*	0.716
MSE	High	1.000*	0.970*	0.863*	1.022*	1.017*	0.908*	0.970*	**0.837***	0.841*	0.907*	0.979*	0.878*	0.945*
	Low	1.000	0.957	0.841*	0.864*	0.986	0.870*	0.863*	0.877*	0.854*	0.862*	**0.822***	0.904*	0.912*
VaR	High	1.000	0.998	**0.854***	0.966	0.946	0.894	0.949	0.868*	0.897	0.896	0.953	0.884	0.932
	Low	1.000	0.983	0.924	0.952	0.960	0.884	0.932	0.880*	0.910	0.887	0.927	**0.878***	0.921

Table 8
Average relative QLIKEs, MSEs, and VaRs for the out-of-sample volatility forecasts at 1-, 5-, 10- and 22-day horizons obtained using alternative approaches, the "insanity filter" of Bollerslev et al. (2016), and a rolling window for the 26 NYSE stocks: Average QLIKE, MSE, and VaR ratios of the HAR to alternative approaches. The lowest ratio in each row is indicated in bold. Numbers in parentheses indicate how many times each approach is included in $\widehat{\mathcal{M}}^*_{90}$.

	HAR	HARQ	RR-HAR	WLS_{RQ}-HAR	log-HAR	LHAR	HAR-RSV	HAR-CJ
				1-day				
QLIKE	1.000$^{(2)}$	1.032$^{(2)}$	1.056$^{(0)}$	0.939$^{(11)}$	**0.934**$^{(13)}$	0.946$^{(21)}$	1.018$^{(2)}$	1.037$^{(0)}$
MSE	1.000$^{(26)}$	0.958$^{(26)}$	0.953$^{(26)}$	0.983$^{(26)}$	0.867$^{(26)}$	**0.859**$^{(26)}$	1.019$^{(26)}$	1.022$^{(26)}$
VaR	1.000$^{(19)}$	1.001$^{(21)}$	0.994$^{(26)}$	0.995$^{(25)}$	**0.993**$^{(25)}$	0.995$^{(24)}$	1.001$^{(19)}$	1.003$^{(21)}$
				5-day				
QLIKE	1.000$^{(11)}$	0.936$^{(23)}$	0.909$^{(19)}$	0.941$^{(25)}$	**0.845**$^{(26)}$	0.884$^{(25)}$	1.000$^{(12)}$	1.051$^{(7)}$
MSE	1.000$^{(26)}$	0.962$^{(26)}$	0.839$^{(26)}$	0.950$^{(26)}$	**0.803**$^{(26)}$	0.855$^{(26)}$	0.993$^{(26)}$	1.023$^{(26)}$
VaR	1.000$^{(0)}$	0.988$^{(0)}$	**0.923**$^{(25)}$	0.988$^{(0)}$	0.932$^{(19)}$	0.930$^{(22)}$	0.998$^{(0)}$	0.996$^{(0)}$
				10-day				
QLIKE	1.000$^{(14)}$	0.934$^{(25)}$	0.876$^{(24)}$	0.960$^{(26)}$	**0.862**$^{(26)}$	0.888$^{(23)}$	0.995$^{(13)}$	1.028$^{(10)}$
MSE	1.000$^{(26)}$	0.981$^{(26)}$	0.874$^{(26)}$	1.013$^{(26)}$	**0.870**$^{(26)}$	0.899$^{(26)}$	0.999$^{(26)}$	1.040$^{(26)}$
VaR	1.000$^{(0)}$	0.994$^{(0)}$	**0.919**$^{(22)}$	0.987$^{(0)}$	0.923$^{(18)}$	0.920$^{(22)}$	0.999$^{(0)}$	0.995$^{(0)}$
				22-day				
QLIKE	1.000$^{(21)}$	0.957$^{(26)}$	**0.903**$^{(26)}$	0.979$^{(26)}$	0.922$^{(26)}$	0.946$^{(23)}$	0.999$^{(21)}$	1.012$^{(21)}$
MSE	1.000$^{(25)}$	0.985$^{(26)}$	**0.897**$^{(26)}$	1.020$^{(26)}$	0.919$^{(26)}$	0.932$^{(26)}$	1.001$^{(25)}$	1.019$^{(25)}$
VaR	1.000$^{(0)}$	0.995$^{(0)}$	0.911$^{(20)}$	0.982$^{(0)}$	0.911$^{(17)}$	**0.908**$^{(23)}$	0.999$^{(0)}$	0.997$^{(0)}$

HAR to alternative approaches. Names ending with -I denote indirect forecast approaches. In all cases with indirect forecasts, one or more of the alternative estimation schemes offer reductions in the loss measures compared to the log-HAR-I estimated by OLS (notably WLS_G for the 5-day horizon and RR for the 10- and 22-day horizons). Some of these reductions are substantial. In sum, the results suggest that indirect forecasts from the HAR model, here applied to logarithmic RV, can benefit at least as much from the alternative estimation schemes (RR, WLS) as direct forecasts do.

We next focus on forecast accuracy. Table 9 shows that the indirect forecasts generally have lower QLIKE, but higher MSE, than the corresponding direct forecasts. However, both approaches are typically included in the MCS for these loss measures. The direct forecasts have lower VaR than the corresponding indirect forecasts in all cases, and only the direct forecast approaches are included in the MCS. Overall, under the alternative estimation schemes, the results indicate that direct forecasts from the log-HAR model perform at least as well as indirect forecasts.

Table 9
Relative QLIKEs, MSEs, and VaRs for the HAR based out-of-sample volatility forecasts at 5-, 10- and 22-day horizons obtained using alternative combination schemes, the "insanity filter" of Bollerslev et al. (2016), and a rolling window for SPX: QLIKE, MSE, and VaR ratios of the direct forecast HAR to alternative approaches. The lowest ratio in each row is indicated in bold. Asterisks indicate approaches included in $\widehat{\mathcal{M}}^{*}_{90}$.

	HAR	log-HAR	log-HAR-I	RR-log-HAR	RR-log-HAR-I	WLS_{RQ}-log-HAR	WLS_{RQ}-log-HAR-I	WLS_{G}-log-HAR	WLS_{G}-log-HAR-I
					5-day				
QLIKE	1.000	0.795	0.786*	0.806	0.788*	0.795	0.791	0.820	**0.779***
MSE	1.000*	0.843*	0.847*	0.852*	0.858*	0.843*	0.851*	0.868*	**0.834***
VaR	1.000	0.932	0.963	**0.923***	0.948	0.935	0.972	0.926*	0.959
					10-day				
QLIKE	1.000	0.741*	0.736*	0.675*	**0.650***	0.745*	0.816*	0.668*	0.804*
MSE	1.000*	0.886*	0.897*	**0.837***	0.844*	0.888*	0.930*	0.844*	0.919*
VaR	1.000	0.908	0.955	**0.898***	0.934	0.910	0.972	0.902	0.956
					22-day				
QLIKE	1.000*	0.812*	0.798*	0.753*	**0.711***	0.810*	0.838*	0.796*	0.783*
MSE	1.000*	0.906*	0.893*	**0.839***	0.843*	0.905*	0.935*	0.880*	0.872*
VaR	1.000	0.889	0.960	**0.873***	0.925	0.892	0.975	0.881	0.951

4. Concluding remarks

This paper explored several, easily implemented, ways to improve the forecasting performance of the standard HAR model. Its main goal was to investigate how the predictive accuracy of the original HAR model depends on choices of estimation scheme, data transformation, and combinations thereof. In an out-of-sample study, covering the S&P 500 index and 26 frequently traded NYSE stocks, it was found that RR and simple WLS schemes can yield substantial improvements to the predictive ability of the HAR model. These simple remedies have the advantage that they can easily be applied directly to the original, linear, HAR model for raw RV and yield an uncomplicated forecast expression. Little evidence in favour of HAR models applied to transformed RV was found. The benefits of replacing OLS with WLS or RR were particularly clear for longer forecast horizons. The possible benefits in predictive accuracy of combinations of different estimation and transformation schemes was also examined. Overall, such combination schemes produced the best performing forecasts, with or without an "insanity filter". The interaction between forecasts from the alternative approaches and the loss functions used for forecast comparison was investigated in terms of over- and under-prediction. It was found that moving away from OLS to WLS or RR typically leads to approaches that are less likely to over-predict, with lower average under-predictions and absolute over-predictions. The results were robust to periods of high and low volatility, and to alternative multi-step ahead forecast schemes. The proposed approaches also performed well compared to some extended HAR models suggested in the literature. These findings provide useful practical insights in the application of the HAR model. Improvements in forecast accuracy can readily be obtained without the need to resort to data beyond publicly available RV and sophisticated extensions of the HAR model.

A number of interesting avenues for future research arise. A more detailed examination of the link between the loss function under which the HAR model parameters are estimated, and the loss function under the intended application of the forecasts would be worthwhile. It would be useful to consider estimating model parameters under a loss function that is coherent with the final application of the forecasts, such as option pricing or Value-at-Risk forecasting. Another natural direction to extend this work is toward the multivariate HAR model (Chiriac and Voev, 2011). However, adapting different estimation schemes and/or transformations in this setting is complicated because of the dimensionality of the problem and the positive definiteness of its associated covariance matrix forecast.

Supplementary material

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.jbankfin.2021.106285.

CRediT authorship contribution statement

Adam Clements: Conceptualization, Methodology, Data curation, Software, Writing – original draft, Writing – review & editing. **Daniel P.A. Preve:** Conceptualization, Methodology, Software, Writing – original draft, Writing – review & editing.

References

Andersen, T., Bollerslev, T., 1998. Answering the skeptics: yes, standard volatility models do provide accurate forecasts. Int Econ Rev (Philadelphia) 39 (4), 885–905.
Andersen, T., Bollerslev, T., Diebold, F., 2007. Roughing it up: including jump components in the measurement, modeling, and forecasting of return volatility. Review of Economics and Statistics 89, 701–720.
Andersen, T., Bollerslev, T., Diebold, F., Labys, P., 2003. Modeling and forecasting realized volatility. Econometrica 71, 579–625.
Barndorff-Nielsen, O., Shephard, N., 2002. Econometric analysis of realized volatility and its use in estimating stochastic volatility models. Journal of Royal Statisitcal Society, Series B 64, 253–280.
Bee, M., Dupuis, D., Trapin, L., 2016. Realizing the extremes: estimation of tail-risk measures from a high-frequency perspective. Journal of Empirical Finance 36, 86–99.
Bekaert, G., Hoerova, M., 2014. The VIX, the variance premium and stock market volatility. J Econom 183 (2), 181–192.
Bollerslev, T., Hood, B., Huss, J., Pedersen, L.H., 2018. Risk everywhere: modeling and managing volatility. Rev Financ Stud 31 (7), 2729–2773.
Bollerslev, T., Patton, A., Quaedvlieg, R., 2016. Exploiting the errors: a simple approach for improved volatility forecasting. J Econom 192, 1–18.
Bollerslev, T., Patton, A., Quaedvlieg, R., 2018. Modeling and forecasting (un)reliable realized covariances for more reliable financial decisions. J Econom 207, 71–91.
Box, G., Cox, D., 1964. An analysis of transformations. Journal of the Royal Statistical Society. Series B (Methodological) 26 (2), 211–252.
Brockwell, P., Davis, R., 1991. Time series: Theory and methods, 2nd Springer Verlag, New York.
Buccheri, G., Corsi, F., 2019. HARK The SHARK: realized volatility modeling with measurement errors and nonlinear dependencies. Journal of Financial Econometrics 1–36.
Cai, J., 1994. A markov model of switching-regime ARCH. Journal of Business & Economic Statistics 12 (3), 309–316.
Chen, X., Ghysels, E., 2011. News–good or bad–and its impact on volatility predictions over multiple horizons. Review of Financial Studies 24 (1), 46–81.
Chevillon, G., Hendry, D., 2005. Non-parametric direct multi-step estimation for forecasting economic processes. Int J Forecast 21 (2), 201–218.
Chiriac, R., Voev, V., 2011. Modelling and forecasting multivariate realised volatility. Journal of Applied Econometrics 26, 922–947.
Cipollini, F., Gallo, G., Otranto, E., 2021. Realized volatility forecasting: robustness to measurement errors. Int J Forecast 37 (1), 44–57.
Corsi, F., 2009. A simple approximate long-memory model of realized volatility. Journal of Financial Econometrics 7 (2), 174–196.

Corsi, F., Mittnik, S., Pigorsch, C., Pigorsch, U., 2008. The volatility of realized volatility. Econom Rev 27 (1–3), 46–78.
Corsi, F., Renò, R., 2012. Discrete-time volatility forecasting with persistent leverage effect and the link with continuous-time volatility modeling. Journal of Business & Economic Statistics 30 (3), 368–380.
Diebold, F., Mariano, R., 1995. Comparing predictive accuracy. Journal of Business and Economics Statistics 13, 253–263.
Engle, R., Rangel, J., 2008. The spline-GARCH model for low-frequency volatility and its global macroeconomic causes. Rev Financ Stud 21, 1187–1222.
Fameliti, S., Skintzi, V., 2020. Predictive ability and economic gains from volatility forecast combinations. J Forecast 39 (2), 200–219.
González-Rivera, G., Lee, T., Mishra, S., 2004. Forecasting volatility: a reality check based on option pricing, utility function, value-at-risk and predictive likelihood. Int J Forecast 20, 629–645.
Hansen, P., Lunde, A., Nason, J., 2003. Choosing the best volatility models: the model confidence set approach. Oxf Bull Econ Stat 65, 839–861.
Hansen, P., Lunde, A., Nason, J., 2011. The model confidence set. Econometrica 79, 453–497.
Marcellino, M., Stock, J., Watson, M., 2006. A comparison of direct and iterated multistep AR methods for forecasting macroeconomic time series. J Econom 135, 499–526.
Muller, U., Dacorogna, M., Dave, R., Olsen, R., Pictet, O., Weizsacker, J., 1997. Volatilities of different time resolutions – analysing the dynamics of market components. Journal of Empirical Finance 4, 213–239.
Patton, A., 2011. Volatility forecast comparison using imperfect volatility proxies. J Econom 160, 246–256.
Patton, A.J., Sheppard, K., 2009. Evaluating Volatility and Correlation Forecasts. In: Andersen, T.G., Davis, R.A., Kreiß, J.P., Mikosch, T. (Eds.), Handbook of Financial Time Series. Springer-Verlag, Berlin.
Patton, A.J., Sheppard, K., 2015. Good volatility, bad volatility: signed jumps and the persistence of volatility. Review of Economics and Statistics 97, 683–697.
Proietti, T., Lütkepohl, H., 2013. Does the box-Cox transformation help in forecasting macroeconomic time series? Int J Forecast 29 (1), 88–99.
Romano, J., Wolf, M., 2017. Resurrecting weighted least squares. J Econom 197, 1–19.
Swanson, N., White, H., 1997. Forecasting economic time series using flexible versus fixed specification and linear versus nonlinear econometric models. Int J Forecast 13 (4), 439–461.
Taylor, N., 2017. Realised variance forecasting under box-Cox transformations. Int J Forecast 33, 770–785.
West, K., 1996. Asymptotic inference about predictive ability. Econometrica 64, 1067–1084.
West, K., 2006. Forecast Evaluation. In: Elliot, G., Granger, C., Timmerman, A. (Eds.), Handbook of Economic Forecasting. Elsevier, Burlington.
Westerlund, J., Narayan, P., 2012. Does the choice of estimator matter when forecasting returns? Journal of Banking and Finance 36, 2632–2640.
Hansen, P., Dumitrescu, E., 2018. Parameter estimation with out-of-sample objective. SSRN working paper, https://dx.doi.org/10.2139/ssrn.3178896.