City Research Online

City, University of London Institutional Repository

Citation: Buccheri, G. & Corsi, F. (2019). HARK the SHARK: Realized Volatility Modeling with Measurement Errors and Nonlinear Dependencies. Journal of Financial Econometrics, 19(4), pp. 614-649. doi: 10.1093/jjfinec/nbz025

This is the accepted version of the paper.

This version of the publication may differ from the final published version.

Permanent repository link: https://openaccess.city.ac.uk/id/eprint/23156/

Link to published version: https://doi.org/10.1093/jjfinec/nbz025

Copyright: City Research Online aims to make research outputs of City, University of London available to a wider audience. Copyright and Moral Rights remain with the author(s) and/or copyright holders. URLs from City Research Online may be freely distributed and linked to.

Reuse: Copies of full items can be used for personal research or study, educational, or not-for-profit purposes without prior permission or charge. Provided that the authors, title and full bibliographic details are credited, a hyperlink and/or URL is given for the original metadata page and the content is not changed in any way.

City Research Online: http://openaccess.city.ac.uk/ publications@city.ac.uk

HARK the SHARK: Realized Volatility Modeling with Measurement Errors and Nonlinear Dependencies

Giuseppe Buccheri[a,*], Fulvio Corsi[b,†]

[a] *Scuola Normale Superiore*

[b] *University of Pisa and City University of London*

This version: April, 2019

Abstract

Despite their effectiveness, linear models for realized variance neglect measurement errors on integrated variance and exhibit several forms of misspecification due to the inherent nonlinear dynamics of volatility. We propose new extensions of the popular approximate long-memory HAR model apt to disentangle these effects and quantify their separate impact on volatility forecasts. By combining the asymptotic theory of the realized variance estimator with the Kalman filter and by introducing time-varying HAR parameters, we build new models that account for: (i) measurement errors (HARK), (ii) nonlinear dependencies (SHAR) and (iii) both measurement errors and nonlinearities (SHARK). The proposed models are simply estimated through standard maximum likelihood methods and are shown, both on simulated and real data, to provide better out-of-sample forecasts compared to standard HAR specifications and other competing approaches.

JEL codes: C22; C53; C58,

Keywords: Realized Volatility, HAR, Measurement Errors, Nonlinear Time Series, Score Driven Models, Kalman Filter.

*Corresponding author: giuseppe.buccheri@sns.it, Scuola Normale Superiore, Piazza dei Cavalieri 7, 56126, Pisa (Italy), tel. 050-509091

†fulvio.corsi@gmail.com.

Electronic copy available at: https://ssrn.com/abstract=3089929

1 Introduction

Estimating and forecasting the volatility of financial markets is a prominent topic in theoretical and applied finance. Andersen and Bollerslev (1998) were the first to advocate the use of realized variance computed from high-frequency data as an accurate proxy of the latent integrated variance. Linear reduced-form specifications for time series of realized variance are widely recognized today as being extremely powerful in predicting financial volatility (Andersen et al. 2003). However, the vast majority of these dynamic specifications ignore two key aspects: (i) realized variance is a noisy estimate of the true integrated variance and (ii) volatility dynamics are highly nonlinear. In particular, the HAR model of Corsi (2009), one of the most popular dynamic specification for realized volatility, suffers from three main forms of misspecification when estimated on realized variance series: (a) biased OLS estimates (Bollerslev et al. 2016a); (b) autocorrelated and highly heteroskedastic residuals (Corsi et al. 2008); (c) time-varying OLS coefficients (Chen et al. 2010). While (a) is related to measurement errors, (b) and (c) are due to nonlinear dependencies, i.e. deviations of the true underlying volatility dynamics from the linear HAR specification.

In this work, we aim to disentangle and quantify the impact of measurement errors and nonlinearities on volatility forecasts provided by the HAR. The strategy adopted here is a step-by-step approach. We first devise different HAR extensions aimed to account separately for each effect. Then, we combine the different approaches in a single model that takes into account all the effects. In doing so, we can quantify the forecast gains resulting from removing each form of misspecification. The specific choice of the HAR model is not restrictive, as the same approach can be applied to any linear specification for realized variance.

As a first step, we write the HAR in a linear state-space representation where the time-varying variance of the measurement error is related to realized quarticity, as prescribed by the asymptotic theory of realized variance (Barndorff-Nielsen and Shephard 2002). The Kalman filter allows to easily estimate bias-corrected HAR parameters and incorporates the effect of measurement errors through a time-varying Kalman gain. We name this model HAR-Kalman (HARK). The HARK corrects HAR forecasts based on the uncertainty with which volatility is measured. It provides more conservative forecasts when current volatility estimates are noisy and, in contrast, generates more responsive forecasts when volatility is measured with a good accuracy.

Compared to the HAR, the $\text{HAR}_{\log}$, i.e. the HAR estimated on $\log(RV_t)$ series, provides a better dynamic specification for realized variance (see e.g. Corsi et al. 2008). However, even on $\log(RV_t)$ series, residuals feature significant heteroskedasticity. In addition, OLS coefficients tend to change significantly when estimating the model on different time windows. These empirical findings reveal that realized variance series are far from being generated by a linear process. We devise an $\text{HAR}_{\log}$ extension that features heteroskedastic errors and time-varying coefficients. In particular, following Creal et al. (2013), we adopt

Electronic copy available at: https://ssrn.com/abstract=3089929

an observation-driven approach where parameters evolve based on the score of the conditional likelihood. The resulting model, the Score-HAR (SHAR), is nonlinear in nature and features i.i.d. standardized residuals when estimated on real data. It is worth underlying that, even if the nonlinearities captured by the SHAR are related to heteroskedasticity and time-varying coefficients, this specification can describe more general forms of nonlinear dependencies. Indeed, Blasques et al. (2014) showed that general nonlinear autoregressive models[1] can be equivalently represented as linear autoregressive models with time-varying parameters. They proved that this formulation is optimal from an information theoretic perspective, provided that parameters are driven by the score of the conditional likelihood. This result is confirmed by our empirical results, which show that the SHAR performs well compared to other nonlinear specifications, like the HARST model of McAleer and Medeiros (2008).

As a final step, we combine together the HARK and the SHAR to obtain a more general model, the SHARK, that accounts for all forms of misspecification. In particular, we let the parameters of the HARK evolve through an observation-driven update scheme. The resulting model is conditionally Gaussian and can be handled by the Kalman filter. As underlined by Harvey (1991), this method allows to introduce nonlinearities into a linear state-space representation while still maintaining a closed form for the likelihood function. Parameters are driven by the score of the conditional likelihood, and thus estimation can be performed as described by Delle Monache et al. (2016). As the HARK, the SHARK can handle measurement errors, but also accounts for heteroskedasticity and time-varying parameters. As the SHAR, the SHARK accounts for nonlinear dependencies, but provides more responsive time-varying parameters, since static parameters are not bias attenuated by measurement errors.

Our empirical analysis, performed on both index future and individual stock data, provides strong evidence that the effects captured by the HARK, SHAR and SHARK are relevant for forecasting volatility. As a matter of fact, the three models are always included in the model confidence set of Hansen et al. (2011), while the $\text{HAR}_{\log}$ and other competing models feature much lower p-values and are excluded in the vast majority of the cases. Interestingly, we found that the SHAR and the SHARK tend to outperform the benchmark models even if the former are estimated on a fixed window and the latter are estimated on a rolling window. Indeed, the SHAR and the SHARK have built-in time-varying HAR coefficients, which automatically adapt to changes in the persistence of volatility. This is true even if the static parameters of the two models are not re-estimated over time.

Comparing the relative forecast gains among our models, we obtain that, from a forecasting perspective, measurement errors are relevant when realized variance is computed at small and intermediate sampling frequencies. We also find that the corresponding forecast gains slightly increase with the forecast horizon.

[1]Nonlinear autoregressive models have the general form $y_t = \phi(Y_{t-1}, \Theta) + u_t$, where $\phi(\cdot)$ is a nonlinear function of past observations $Y_{t-1} = \{y_{t-1}, y_{t-2}, \dots\}$, Θ is a set of parameters and u_t is a zero-mean sequence of independent innovations.

Electronic copy available at: https://ssrn.com/abstract=3089929

The impact of heteroskedasticity is always relevant, especially at long forecast horizons. Finally, time-varying parameters provide statistically significant forecast gains that are independent from the sampling frequency and slightly increase with the forecast horizon.

The problem of taking into account the effect of measurement errors on realized variance forecasts was recently tackled by Bollerslev et al. (2016a), who devised the HARQ model. The main idea is to augment the HAR with a term proportional to lagged RV_t's that also depends on realized quarticity. The new term adjusts HAR forecasts based on the current level of uncertainty on realized variance. Our approach differs in several aspects. First, assuming an HAR plus noise as a DGP for the realized variance, the Kalman filter provides MVLUE estimates (see e.g. Durbin and Koopman 2012) and corrections due to measurement errors are independent from the level of heteroskedasticity of the noise. Second, forecasting gains provided by the HARQ may in principle be imputable to potential nonlinear dependencies captured by the correction term, as recently pointed out by Cipollini et al. (2017), or result from a mixture of measurement errors and nonlinearity. Instead, our approach allows to quantify the separate effect of estimation errors, after disentangling it from nonlinearities.

Our work is also related to that of Asai et al. (2012), who employed a state-space representation to account for measurement errors. However, the error variance in their work is homoskedastic and is not related to realized quarticity. Bekierman and Manner (2018) proposed a modification of the HARQ model by allowing the daily coefficient of the HAR to be driven by a latent Gaussian process. This specification is nested into our SHAR model, which has also time-varying weekly and monthly coefficients and heteroskedastic innovations. The HAR model has been extended into several directions. For instance, McAleer and Medeiros (2008) proposed a smooth transition HAR specification to capture different volatility regimes. Andersen et al. (2007) and Patton and Sheppard (2015) showed how to improve predictability by separating continuous and jump components, and by separating the volatility of positive and negative returns, respectively. Corsi and Renò (2012) included the effect of leverage. Our formulatión can easily be extended to include these new covariates in the state-space representation. For instance, in the online appendix, we show how to include leverage in the HARK, SHAR and SHARK models. Finally, Chen et al. (2018) proposed a non-parametric time-varying HAR specification. Other examples of nonlinear models for realized variance are the class of MEM models developed by Engle (2002), Engle and Gallo (2006), Cipollini et al. (2017).

All our three HAR extensions are formulated on $\log(RV_t)$ series. Therefore, when forecasting, one needs to take into account the Jensen inequality. We devise simple methods to compute bias-corrected one-step and multi-steps ahead forecasts.

The rest of the paper is organized as follows. Section 2 introduces the main concepts and describes all the details related to the estimation of our three HAR extensions; in Section 3, we provide empirical evidence

Electronic copy available at: https://ssrn.com/abstract=3089929

on the relevance of the effects captured by our models; finally, Section 4 concludes. Complementary results and proofs are reported in the online appendix.

2 Theoretical framework

2.1 Asymptotic theory of realized variance estimation

Let $\mathcal{P}_s$ denote the asset price at time s. We assume that the log-price $X_s = \log(\mathcal{P}_s)$ evolves according to a Brownian semimartingale process:

$$dX_s = \mu_s ds + \sigma_s dW_s \tag{1}$$

where μ_s and σ_s are drift and instantaneous volatility processes satisfying the usual assumptions, and W_s is a Wiener process. The day t integrated variance of X_s is defined as:

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

where the unit time interval corresponds to one trading day. Note that IV_t turns out to be equal to the quadratic variation of X_s computed on the interval $[t-1, t]$ (see e.g. Protter 1992).

The day t realized variance is defined as:

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

where the intraday returns $r_{i,t} = X_{t-1+i\Delta} - X_{t-1+(i-1)\Delta}$, $i = 1, \ldots, M$ are computed on M intraday time intervals of length $\Delta = 1/M$. The econometric theory of Barndorff-Nielsen and Shephard (2002) suggests that, as the intraday period Δ goes to zero, the estimation error is mixed normal distributed:

$$RV_t = IV_t + \epsilon_t, \quad \epsilon_t \sim \text{MN}(0, 2\Delta IQ_t) \tag{4}$$

where $IQ_t = \int_{t-1}^{t} \sigma^4(s) ds$ is the integrated quarticity of the underlying semimartingale process. IQ_t can consistently be estimated using the realized quarticity estimator:

$$RQ_t = \frac{M}{3} \sum_{i=1}^{M} r_{i,t}^4 \tag{5}$$

For later convenience, we also report the asymptotic distribution of $\log(RV_t)$. Using the delta method, it is immediate to see that, as $\Delta \to 0$:

$$\log(RV_t) = \log(IV_t) + \xi_t, \quad \xi_t \sim \text{MN}\left(0, 2\Delta \frac{IQ_t}{IV_t^2}\right) \tag{6}$$

A consistent estimator V_t of the variance of ξ_t can be obtained by replacing IV_t and IQ_t with their consistent estimators (3), (5), namely:

$$V_t = \frac{2}{3} \frac{\sum_{i=1}^{M} r_{i,t}^4}{(\sum_{i=1}^{M} r_{i,t}^2)^2} \tag{7}$$

Electronic copy available at: https://ssrn.com/abstract=3089929

As pointed out by Barndorff-Nielsen and Shephard (2002), this finite sample approximation is quite accurate even at moderate values of M, where the approximation (4) is instead less reliable. Note that jumps and microstructure effects have not been considered. In order take them into account, it is sufficient to replace RV_t and RQ_t with robust estimators. However, our general approach remains unchanged.

2.2 Reduced-form models for volatility estimation and forecasting

If one neglects the measurement error term in eq. (4), then $RV_t = IV_t$ and the integrated variance is observable. This assumption is at the basis of the reduced-form specifications that have been employed in last years to model and forecast volatility using time series of realized variance.

One of the most popular linear specification is the approximate long-memory HAR model of Corsi (2009). The HAR reads[2]:

$$RV_{t+1} = \beta_0 + \beta_1 RV_t + \beta_2 RV_{t-1|t-4} + \beta_3 RV_{t-5|t-21} + \eta_{t+1} \tag{8}$$

where η_{t+1} are i.i.d. innovations and $RV_{t_1|t_2}$ denotes the average of daily RV's from day t_1 to day t_2. Formally, the HAR is not a long-memory process. However, the aggregation of volatilities at short and long time-scales leads to a slowly decaying autocorrelation function that closely resembles the one observed on real financial data. The model can be simply estimated by OLS and provides out-of-sample forecasts that have been proved to perform as well as those of long-memory ARFIMA models (Corsi 2009). Given the simplicity and effectiveness of the HAR, we use it as a basis for our modeling framework and empirical work. However, any linear specification for RV_t is susceptible of treatment within this framework.

2.3 Measurement errors

The main effect of the measurement error term in eq. (4) is the well-known attenuation bias of OLS coefficients in presence of latent regressors (see e.g. Wansbeek and Meijer 2000). As a consequence, the estimated model features lower persistence and volatility forecasts are less accurate. Recently, there was an increased interest in examining the effect of measurement errors on HAR forecasts. Bollerslev et al. (2016a) suggested to augment the HAR with a term depending on realized quarticity. The latter underrates RV forecasts in case estimation errors are large, and generates more responsive forecasts when errors are small. Their HARQ model reads:

$$RV_{t+1} = \beta_0 + (\beta_1 + \beta_{1Q} RQ_t^{1/2}) RV_t + \beta_2 RV_{t-1|t-4} + \beta_3 RV_{t-5|t-21} + \eta_{t+1} \tag{9}$$

If $\beta_{1Q} < 0$, the term $\beta_{1Q} RQ_t^{1/2}$ corrects RV_{t+1} based on the degree of uncertainty with which RV_t is measured. Similar correction terms can be included in the weekly and monthly regressors.

[2]Note that this HAR specification is identical to that in Patton and Sheppard (2015) and Corsi et al. (2013). We use it throughout the paper.

Electronic copy available at: https://ssrn.com/abstract=3089929

In order to examine the effect of the correction provided by the HARQ, we simulate IV_t using an HAR specification and contaminate the simulated observations with a white noise term:

$$RV_t = IV_t + \epsilon_t \tag{10}$$

$$IV_{t+1} = \beta_0 + \beta_1 IV_t + \beta_2 IV_{t-1|t-4} + \beta_3 IV_{t-5|t-21} + \eta_{t+1} \tag{11}$$

where $\epsilon_t \sim \text{NID}(0, h_t)$ and $\eta_{t+1} \sim \text{NID}(0, q)$ are measurement and model disturbances with $\mathbb{E}[\epsilon_t, \eta_t] = \rho$. Observations are simulated by first sampling IV_t, $t = 1, \ldots, 22$, from a lognormal distribution with mean and variance given by the unconditional mean and variance of the process (11), and then iteratively applying eq. (10) and (11). Since the HARQ relies on having an heteroskedastic measurement error, we model dynamically h_t through a GARCH process:

$$z_t = \sqrt{h_t}\zeta_t \tag{12}$$

$$h_{t+1} = \omega + \gamma z_t^2 + \nu h_t \tag{13}$$

where $\zeta_t \sim \text{NID}(0, 1)$. We define the signal-to-noise ratio as $\delta = q(1 - \gamma - \nu)/\omega$, and in this example we set $\rho = 0.4$, $q = 0.1$, $\gamma = 0.01$, $\rho = 0.95$ and $\omega = q(1 - \gamma - \rho)$, such that $\delta = 1$. The remaining parameters are set as $\beta_0 = 1$, $\beta_1 = 0.5$, $\beta_2 = 0.2$, $\beta_3 = 0.1$.

After generating $T = 1000$ observations for $N = 1000$ Monte-Carlo repetitions, the simulated RV_t and h_t are used to estimate[3] the HAR and HARQ models. In Figure (1), we plot kernel density estimates of $z_i = \hat{\theta}_i - \theta_i$, where θ_i, $\hat{\theta}_i$, $i = 1, \ldots, 4$, denote the parameters of the HAR specification in eq. (11) and their estimates obtained using the HAR and HARQ models. In both cases, the coefficient β_1 of RV_t exhibits a strong negative bias. However, the bias is lower in the HARQ, which in turn is more persistent than the standard HAR model. The other two coefficients are instead less biased, since measurement errors are smoothed out when averaging lagged RV_t 's.

Note that the estimated HARQ coefficient β_{1Q} depends on the level of heteroskedasticity of ϵ_t. If h_t is homoskedastic, β_{1Q} is not identifiable, even in presence of large measurement error variance. If h_t is time-varying, larger variations are related to larger estimates of β_1, which lead to better forecasts over the standard HAR model.

2.4 The HARK model

The Kalman filter provides a natural way of recovering consistent and unbiased estimates of the HAR coefficients when observations of IV_t are contaminated by noise. Let $\mathcal{F}_t$ be the σ-field generated by RV_t

[3]On real data h_t is not available and is replaced by the RQ_t estimator in eq. (5).

Electronic copy available at: https://ssrn.com/abstract=3089929

and RQ_t. Let us consider the following linear state-space representation:

$$RV_t = Z\alpha_t + \epsilon_t, \quad \epsilon_t \sim \text{NID}(0, h_t) \tag{14}$$

$$\alpha_{t+1} = c + T\alpha_t + \eta_t, \quad \eta_t \sim \text{NID}(0, Q) \tag{15}$$

where we have introduced the 22×1 state vector and the 22×22 transition matrix:

$$\alpha_{t+1} = \begin{pmatrix} IV_{t+1} \\ IV_t \\ \vdots \\ IV_{t-20} \end{pmatrix}, \quad T = \begin{pmatrix} \beta_1, & \frac{1}{4}\beta_2 & \overbrace{\cdots}^{4\,terms} & \frac{1}{4}\beta_2, & \frac{1}{17}\beta_3 & \overbrace{\cdots}^{17\,terms} & \frac{1}{17}\beta_3 \\ 1 & 0 & \cdots & \cdots & \cdots & 0 & 0 \\ 0 & \ddots & & & & \vdots & \vdots \\ \vdots & & \ddots & & & \vdots & \vdots \\ \vdots & & & \ddots & & \vdots & \vdots \\ \vdots & & & & \ddots & \vdots & \vdots \\ 0 & \cdots & \cdots & \cdots & \cdots & 1 & 0 \end{pmatrix} \tag{16}$$

together with the 22×1 vector of constants and the 22×22 covariance matrix:

$$c = \begin{pmatrix} \beta_0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \quad Q = \begin{pmatrix} q & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 0 \end{pmatrix} \tag{17}$$

The 22-dimensional row vector $Z = (1, 0 \dots, 0)$ selects the first element of α_{t+1}. The variance of the measurement error h_t is assumed to be measurable with respect to the information set $\mathcal{F}_t$.

Model (14), (15) is a linear Gaussian state-space representation that can be estimated through standard maximum likelihood methods using the Kalman filter (see e.g. Durbin and Koopman 2012). The latter allows to recursively compute the conditional mean $a_{t+1} = \text{E}[\alpha_{t+1}|\mathcal{F}_t]$ and the conditional variance $P_{t+1} = \text{Var}[\alpha_{t+1}|\mathcal{F}_t]$ of the state vector α_t. The log-likelihood can be computed as:

$$\log L(RV_1, \dots, RV_T|\Phi) = -\frac{22}{2}T\log(2\pi) - \frac{1}{2}\sum_{t=1}^{T}\left(\log|F_t| + v_t' F_t^{-1} v_t\right) \tag{18}$$

where $v_t = RV_t - Za_t$ is the prediction error, $F_t = ZP_tZ' + h_t$ is the corresponding covariance matrix and $\Phi = \{\beta_0, \beta_1, \beta_2, \beta_3, q\}$ is the set of parameters of the model. In Section (D) in the online appendix, we report the Kalman filter recursions to compute a_t and F_t. Parameters are estimated by maximizing the log-likelihood through a quasi-Newton algorithm:

$$\hat{\Phi} = \underset{\Phi}{\text{argmax}} \log L(RV_1, \dots, RV_T|\Phi) \tag{19}$$

In order to select a proxy for the variance h_t of the measurement error term in eq. (14), we rely on the asymptotic theory of Barndorff-Nielsen and Shephard (2002) in eq. (4) and use the realized quarticity

Electronic copy available at: https://ssrn.com/abstract=3089929

estimator in eq. (5). In Section (3.4), we will examine more flexible specifications for h_t to test deviations from this assumption.

The one-step-ahead prediction in the Kalman filter is given by $a_{t+1} = c + Ta_t + K_t v_t$, where $K_t = TP_t Z' F_t^{-1}$ is the Kalman gain. If the error variance h_t is large, meaning that RV_t is estimated with large uncertainty, the predicted variance is penalized by a small gain. In contrast, if h_t is small, meaning that RV_t is estimated with great accuracy, the predicted variance turns out to be more responsive to RV_t. We name this model HAR-Kalman (HARK).

Figure (2) shows kernel density estimates of the pivotal HARK statistics obtained using the same data generating process of Section (2.3). As done with the HARQ in the previous experiment, we use the true measurement error variance h_t to estimate the HARK. The estimates provided by the HARK are unbiased and distributed according to a normal, as predicted by the econometric theory on linear state-space models. The estimation of the model is very fast, since system matrices are sparse.

It is important to examine whether improvements in parameter estimates translate into better out-of-sample forecasts. For different values of the signal-to-noise ratio δ, we generate $N = 250$ Monte-Carlo realizations of model (10), (11) with $T = 2000$ observations. The latter 1000 observations are predicted using a rolling window of 1000 observations. In Figure (3), we show the out-of-sample MSE of HARQ and HARK models relative to the MSE of the HAR model, for different values of the signal-to-noise ratio δ. As expected, there are large forecast gains when δ is small, since the observations of the underlying volatility process are mainly dominated by noise. However, the MSE provided by the HARK is always lower than that of the HARQ. This is not surprising, as the HARQ model is misspecified when the true DGP is the one in eq. (10), (11), while the HARK provides the correct specification[4].

2.5 The SHAR model

The DGP (10), (11) used in the previous analysis takes into account the effect of measurement errors when forecasting IV_t, but is characterized by a linear transition equation. In order to understand whether a linear specification fits well real data, we consider high-frequency futures prices of the S&P 500 market index and compute daily series of realized variance by summing squares of 5-min returns. The resulting time-series is shown in Figure (4) and includes 4259 observations, from 03-01-1995 to 21-06-2013.

It is known that the $\text{HAR}_{\log}$, i.e. the HAR estimated on $\log(RV_t)$ series, features less heteroskedastic residuals (see e.g. Corsi et al. 2008). This is evident from figure (5), which shows the OLS residuals of

[4]Note that the standard linear Gaussian representation in eq. (14), (15) assumes that the two disturbances ϵ_t and η_t are uncorrelated. The Kalman filter can easily be extended to include correlated measurement and innovation disturbances, see e.g. Simon (2006). However, this modification does not lead to significant out-of-sample forecast gains on real data. For simplicity, we thus consider the standard linear-Gaussian representation with uncorrelated disturbances.

Electronic copy available at: https://ssrn.com/abstract=3089929

HAR and $\text{HAR}_{\log}$, and from figure (6), which shows the sample autocorrelation function of residuals and squared residuals of the two models. Compared to the HAR, the residuals of the $\text{HAR}_{\log}$ are uncorrelated, while squared residuals have weaker serial correlation.

As a consequence, being the HARK based on a linear state-space representation, it is convenient to formulate it on $\log RV_t$, rather then on RV_t. The only modification to the formulation in Section (2.4) is that we now rely on the asymptotic distribution of $\log RV_t$ and, consequently, the measurement error variance h_t is computed using eq. (6) and (7). It is useful to investigate whether the weaker misspecification of the $\text{HAR}_{\log}$ model translates into superior out-of-sample forecasts compared to the HAR model on RV_t. In the empirical application in Section (3), we show that the $\text{HAR}_{\log}$ outperforms the HAR, with forecast gains that are statistically significant for all the time-series under investigation. A similar result is found in the simulation study in Section (A) of the online appendix, where we consider a DGP that closely resembles real financial time-series.

We can write the $\text{HAR}_{\log}$ model as:

$$RV_{t+1}^l = \beta_0 + \beta_1 RV_t^l + \beta_2 RV_{t-1|t-4}^l + \beta_3 RV_{t-5|t-21}^l + \eta_{t+1}, \quad \eta_{t+1} \sim \text{NID}(0, q) \tag{20}$$

where $RV_t^l = \log(RV_t)$, and $RV_{t-1|t-4}$, $RV_{t-5|t-21}$ are built as in the HAR, but are computed through $\log(RV_t)$. The heteroskedasticity featured by the $\text{HAR}_{\log}$ suggests that this residual misspecification can be removed by introducing a time-varying conditional variance for the innovations η_t. The $\text{HAR}_{\log}$ coefficients are also unstable over time. Figure (7) shows the dynamics of $\beta_0, \beta_1, \beta_2, \beta_3$ obtained by estimating the $\text{HAR}_{\log}$ on daily realized variances of S&P500 future prices using a rolling window of 1000 observations. Variations over time of the estimated coefficients are significant and suggest that volatility forecasts might be improved by modeling dynamically the $\text{HAR}_{\log}$ coefficients.

We consider the following $\text{HAR}_{\log}$ model with time-varying parameters:

$$RV_{t+1}^l = \beta_{0,t+1} + \beta_{1,t+1} RV_t^l + \beta_{2,t+1} RV_{t-1|t-4}^l + \beta_{3,t+1} RV_{t-5|t-21}^l + \eta_{t+1}, \quad \eta_{t+1} \sim \text{NID}(0, q_{t+1}) \tag{21}$$

Score-driven models, introduced by Creal et al. (2013) and Harvey (2013), provide a useful methodological framework to model time-varying parameters. In this class of observation-driven models, parameters change over time based on the score of the conditional density. We define the vector of time-varying parameters at time t as:

$$f_t = \left(\beta_{0,t}, \beta_{1,t}, \beta_{2,t}, \beta_{3,t}, \log q_t\right)'$$

The next value f_{t+1} is determined by the following update rule:

$$f_{t+1} = \omega + A s_t + B f_t \tag{22}$$

Electronic copy available at: https://ssrn.com/abstract=3089929

where s_t is the scaled score vector:

$$s_t = (\mathcal{I}_{t|t-1})^{-1}\nabla_t, \quad \nabla_t = \left[\frac{\partial \log p(RV_t^l|f_t, \mathcal{B}_{t-1}, \Theta)}{\partial f_t'}\right]', \quad \mathcal{I}_{t|t-1} = \mathrm{E}[\nabla_t \nabla_t']$$

and $\mathcal{B}_t$ denotes the information set generated by current and past observations of RV_t. The inverse of the information matrix $\mathcal{I}_{t|t-1}$ is introduced to correct for the curvature of the likelihood function. The vector $\Theta = \{\omega, \mathrm{vec}(A), \mathrm{vec}(B)\}$ denotes the set of the static parameters of the model. The conditional log-likelihood is given by:

$$\log p(RV_t^l|f_t, \mathcal{B}_{t-1}, \Theta) = -\frac{1}{2}\left(\log q_t + \frac{(RV_t^l - \mu_{t|t-1})^2}{q_t}\right) \tag{23}$$

where $\mu_{t|t-1}$ is the conditional mean:

$$\mu_{t|t-1} = \mathrm{E}[RV_t^l|f_t, \mathcal{B}_{t-1}] = \beta_{0,t} + \beta_{1,t}RV_{t-1}^l + \beta_{2,t}RV_{t-2|t-5}^l + \beta_{3,t}RV_{t-6|t-22}^l \tag{24}$$

The static parameters Θ are estimated by numerically optimizing the log-likelihood function:

$$\hat{\Theta} = \underset{\Theta}{\mathrm{argmax}} \log L(RV_1^l, \ldots, RV_T^l|\Theta) \tag{25}$$

where $\log L(RV_1^l, \ldots, RV_T^l|\Theta)$ is computed as:

$$\log L(RV_1^l, \ldots, RV_T^l|\Theta) = \sum_{t=1}^{T} \log p(RV_t^l|f_t, \mathcal{B}_{t-1}, \Theta) \tag{26}$$

We denote model (21), equipped with the update rule in eq. (22), as Score-HAR (SHAR) model. In Section (E) in the online appendix, we will recover formulas for ∇_t and $\mathcal{I}_{t|t-1}$. Note that, if we set $f_t = \log q_t$, i.e. if we neglect dynamics in the HAR coefficients, we obtain the heteroskedastic HAR model of Corsi et al. (2008) as a particular case of the SHAR model.

As shown by Blasques et al. (2014), autoregressive models with time-varying coefficients can be employed as alternative representations for general nonlinear autoregressive models. This result implies that the SHAR can capture more general forms of nonlinear dependencies. Indeed, in the empirical analysis in Section (3), we show that the SHAR performs well compared to the HARST model of McAleer and Medeiros (2008), which is a nonlinear smooth transition HAR specification with multiple volatility regimes.

In order to examine whether the SHAR is able to remove the misspecification of the $\mathrm{HAR}_{\log}$, we look at the standardized residuals $\hat{\eta}_t = (RV_t^l - \mu_{t|t-1})/\sqrt{q_t}$ obtained by estimating the model on the same sample used in the previous analysis. Figure (8) shows sample autocorrelations of $\hat{\eta}_t$ and $\hat{\eta}_t^2$, while Figure (9) shows kernel density estimates of $\hat{\eta}_t$. The weak heteroskedasticity of $\mathrm{HAR}_{\log}$ residuals in Figure (6) has now disappeared and $\hat{\eta}_t$ is i.i.d. Slight deviations from normality in the tails are imputable to the huge spikes observed during the 2008-2009. Similar results are obtained by estimating the SHAR on individual stock data.

Electronic copy available at: https://ssrn.com/abstract=3089929

2.6 The SHARK model

The HARK accounts for measurement errors when estimating and forecasting volatility with an $\text{HAR}_{\log}$ specification. However, the $\text{HAR}_{\log}$ exhibits heteroskedastic residuals and time-varying coefficients. As done with the SHAR, we can introduce time-varying parameters in the HARK in order to correct for its residual misspecification. The new model reads:

$$RV_t^l = Z\alpha_t + \epsilon_t, \quad \epsilon_t \sim \text{NID}(0, h_t) \tag{27}$$

$$\alpha_{t+1} = c_t + T_t\alpha_t + \eta_t, \quad \eta_t \sim \text{NID}(0, Q_t) \tag{28}$$

where the system matrices have the same form as in eq. (16), (17), but parameters $\beta_{0,t}, \beta_{1,t}, \beta_{2,t}, \beta_{3,t}, q_t$ are now time-varying.

Score-driven models can be employed to model time-varying parameters in a linear Gaussian state-space representation. Indeed, if one assumes that the time-varying system matrices c_t, T_t and Q_t are known given past observations, the model is conditionally Gaussian and can still be handled by the Kalman filter. By doing so, the model features nonlinear dynamics, but the state-space representation maintains a closed form for the likelihood function. Creal et al. (2008) and Delle Monache et al. (2016) showed that, by combining the Kalman filter and the score-driven filter, one obtains a unified filter for both the latent state vector and the time-varying parameters. As in the SHAR, the latter are given by

$$f_t = \Big(\beta_{0,t}, \beta_{1,t}, \beta_{2,t}, \beta_{3,t}, \log q_t\Big)' \tag{29}$$

and follow the usual update rule:

$$f_{t+1} = \zeta + Cs_t + Df_t \tag{30}$$

where $s_t = (\mathcal{I}_{t|t-1})^{-1}\nabla_t$. As shown by Delle Monache et al. (2016), ∇_t and $\mathcal{I}_{t|t-1}$ can be computed as[5]:

$$\nabla_t = -\frac{1}{2}\left[\dot{F}_t'(\mathbb{I}_{n_t} \otimes F_t^{-1})\text{vec}(\mathbb{I}_{n_t} - v_t v_t' F_t^{-1}) + 2\dot{v}_t' F_t^{-1} v_t\right] \tag{31}$$

$$\mathcal{I}_{t|t-1} = \frac{1}{2}\left[\dot{F}_t'(F_t^{-1} \otimes F_t^{-1})\dot{F}_t + 2\dot{v}_t' F_t^{-1}\dot{v}_t\right] \tag{32}$$

where v_t and F_t are the Kalman filter prediction error and its covariance matrix. The two quantities v_t and F_t, are computed as an output of the Kalman filter with time-varying parameters. Instead, $\dot{v}_t$ and $\dot{F}_t$ denote derivatives of v_t and F_t with respect to f_t, namely $\dot{v}_t = \partial v_t/\partial f_t'$ and $\dot{F}_t = \partial \text{vec}(F_t)/\partial f_t'$. They can be computed through a parallel set of recursions that is reported in Section (G) in the online appendix.

The log-likelihood is conveniently computed in the prediction-error decomposition:

$$\log L(RV_1^l, \ldots, RV_T^l|\Omega) = -\frac{22}{2}T\log(2\pi) - \frac{1}{2}\sum_{t=1}^{T}\left(\log|F_t| + v_t' F_t^{-1} v_t\right) \tag{33}$$

[5]The notation is described in detail in the online appendix.

Electronic copy available at: https://ssrn.com/abstract=3089929

where $\Omega = \{\zeta, \text{vec}(C), \text{vec}(D)\}$ denotes the set of static parameters of the model. The time-varying parameters are updated at each time step using eq. (30) and computing ∇_t and $\mathcal{I}_{t|t-1}$ through eq. (31), (32). The static parameters are estimated by optimizing numerically the log-likelihood with a quasi-Newton method:

$$\hat{\Omega} = \underset{\Omega}{\text{argmax}} \log L(RV_1^l, \dots, RV_T^l | \Omega) \tag{34}$$

We denote model (27), (28), equipped with the update rule in eq. (30), as Score-HAR-Kalman (SHARK). As we will see in our empirical study in Section (3), compared to the SHAR, the SHARK provides more responsive time-varying parameters, as they are not affected by the attenuation bias due to measurement errors.

2.7 Forecast

In this Section, we provide guidelines for computing one-step and multi-step-ahead forecasts from our models. Except for the HAR, all other models are estimated on $\log(RV_t)$ series. Thus, when computing forecasts of RV_t, it is essential to take into account the bias generated by the logarithm transformation.

2.7.1 One-step-ahead

The one-step-ahead forecast of the $\text{HAR}_{\log}$ model is simply computed using the moment generating function of the normal distribution:

$$\text{E}_t[\exp(RV_{t+1}^l)] = \exp\left(\beta_0 + \beta_1 RV_t^l + \beta_2 RV_{t-1|t-4}^l + \beta_3 RV_{t-5|t-21}^l + \frac{q^2}{2}\right) \tag{35}$$

In the HARK, from the theory of linear Gaussian state-space models, we know that $\alpha_{t+1}|\mathcal{F}_t \sim \text{N}(a_{t+1}, P_{t+1})$. Thus, using the moment generating function of the multivariate normal distribution, we have:

$$\text{E}_t[\exp(Z\alpha_{t+1})] = \exp\left(Za_{t+1} + \frac{1}{2} Z P_{t+1} Z'\right) \tag{36}$$

Since the time-varying parameters in the SHAR and SHARK are one-step ahead predictable, we obtain forecast formulas similar to those in eq. (35) and (36), respectively. In particular, in the case of the SHAR, we have:

$$\text{E}_t[\exp(RV_{t+1}^l)] = \exp\left(\beta_{0,t+1} + \beta_{1,t+1} RV_t^l + \beta_{2,t+1} RV_{t-1|t-4}^l + \beta_{3,t+1} RV_{t-5|t-21}^l + \frac{q_{t+1}^2}{2}\right) \tag{37}$$

In the case of the SHARK, we obtain the same one-step-ahead forecast of the HARK in eq. (36), but a_{t+1} and P_{t+1} are computed with the Kalman filter with time-varying parameters given in the online Appendix F.

Electronic copy available at: https://ssrn.com/abstract=3089929

2.7.2 Multi-step-ahead

We provide closed form bias-corrected expressions for multi-step ahead forecasts of $\text{HAR}_{\log}$ and HARK models. In the SHAR and SHARK models, the presence of time-varying parameters does not allow for similar closed form expressions. In this case, as suggested by Harvey (2013) and Creal et al. (2014), we evaluate the conditional mean of the predictive density through simulations. This is easily done by simulating recursively eq. (21), (22) for the SHAR and eq. (27), (28), (30) for the SHARK.

As done with the HARK, we write the $\text{HAR}_{\log}$ model in a vectorial representation by introducing the 22-dimensional column vector $\overline{RV}_t^l = (RV_{t+1}^l, RV_t^l, \cdots, RV_{t-21}^l)'$. Eq. (20) can thus be written as:

$$\overline{RV}_{t+1}^l = c_l + T_l \overline{RV}_t^l + \overline{\eta}_t, \quad \overline{\eta}_t \sim \text{NID}(0, Q_l) \tag{38}$$

where c_l, T_l and Q_l have the same structure as c, T and Q in Section (2.4), but contain the parameters of the $\text{HAR}_{\log}$ model. In the online appendix, we prove the following two propositions:

Proposition 1. *The j-th step-ahead forecast of the HAR_{log} model is given by:*

$$E_t[\exp(Z\overline{RV}_{t+j})] = \exp\left\{ Z\left[(\mathbb{I}_n + \cdots + T_l^{j-1})c_l + T_l^j \overline{RV}_t^l\right] + \frac{1}{2} Z\left[Q + \cdots + T^{j-1} Q (T^{j-1})'\right] Z' \right\}$$

Proposition 2. *The j-th step-ahead forecast of the HARK model is given by:*

$$E_t[\exp(Z\alpha_{t+j})] = \exp\left\{ Z\left[(\mathbb{I}_n + \cdots + T^{j-2})c + T^{j-1} a_{t+1}\right] + \right.$$
$$\left. \frac{1}{2} Z\left[T^{j-1} P_{t+1} (T^{j-1})' + Q + \cdots + T^{j-2} Q (T^{j-2})'\right] Z' \right\}$$

3 Empirical evidence

In order to test the proposed models on real time series, we use high-frequency data of both S&P500 market index future and individual stock prices. The S&P500 sample is the same used in Section (2.5). It includes 4259 business days, from 03-01-1995 to 21-06-2013. Equity data are provided by Thomson Reuters and comprise 18 frequently traded NYSE stocks. The sample goes from 03-01-2006 to 31-12-2014, spanning a total of 2250 business days. Both datasets contain one-second transaction prices from 9:30 to 16:00.

Since one of the main purposes of our analysis is to assess the effect of measurement errors, we compute realized variance at different sampling frequencies. We set $M = 39, 78, 390$, corresponding to sampling 10, 5, 1-minute returns[6]. Summary statistics of the series of realized variances are reported in the online appendix.

[6] On S&P500 data, the average realized variance is approximately constant over different sampling frequencies. For $M = 39, 78, 390$, it is equal to $0.90, 0.92, 0.89$, respectively, suggesting that microstructure effects are small. Therefore, the standard realized variance estimator provides unbiased estimates of IV_t for all the three sampling frequencies. Similar considerations hold for individual stock data.

Electronic copy available at: https://ssrn.com/abstract=3089929

We perform the out-of-sample analysis using both a static window (SW) and a rolling window (RW). The length of the rolling window is 2000 days in both datasets. In the case of S&P500 data, we predict the last 2259 observations, from December 2002 to June 2013. In the case of NYSE stock data, we predict the last 250 observations, corresponding to all the business days of 2014. The in-sample analysis is performed on the last 2259 observations on S&P500 data. In the case of NYSE stock data, the last 250 observations are not sufficient to obtain a good inference. We thus perform the in-sample analysis on the last 1000 observations.

We used the following parsimonious specification for the dynamic equations (22), (30) in the SHAR and SHARK:

$$f_{t+1} = f_t + As_t \tag{39}$$

$$f_{t+1} = f_t + Cs_t \tag{40}$$

where A and C are 5×5 diagonal matrices. The static parameters in the SHAR are the five diagonal elements of A, that we denote by a_1, a_2, a_3, a_4, a_5. Similarly, the static parameters in the SHARK are the five diagonal elements of C, that we denote by c_1, c_2, c_3, c_4, c_5. More flexible specifications are possible, but they do not lead to superior out-of-sample forecasts.

The HARQ may provide negative RV_t estimates. In that case, as suggested by Bollerslev et al. (2016b), we replace the negative estimates with the average of past realized variances. As a benchmark, we include an ARFIMA(1,d,1) with AR and MA parameters denoted by ψ and χ, respectively. We also include the smooth-transition HAR (HARST) model of McAleer and Medeiros (2008) and the HAR$^{\Delta J^{2\pm}}$ model of Patton and Sheppard (2015). The latter separates the daily volatility into a continuous component given by the realized bipower variation, and two signed jump components. The ARFIMA and the HARST are estimated on $\log(RV_t)$, while the HAR$^{\Delta J^{2\pm}}$ is estimated on RV_t. Indeed, in the latter case we did not find a statistical significant forecast gain when taking the logarithm. As suggested by McAleer and Medeiros (2008), we compute the forecast as a simple model average of the linear HAR$_{\log}$ and the HARST. Moreover, we use daily log-returns as a transition variable.

3.1 In-sample analysis

In tables (1)-(3), we report parameter estimates of HAR, HARQ, HAR$_{\log}$, HARK, SHAR, SHARK, ARFIMA, HARST, HAR$^{\Delta J^{2\pm}}$ models on the whole S&P500 dataset, excluding the first 500 observations. This pre-sample of 500 observations is used for initializing the time-varying parameters in the SHAR and SHARK models. Specifically, The HAR$_{\log}$ and the HARK are first estimated on the pre-sample, and the estimated HAR coefficients are used as starting values for the time-varying parameters of the SHAR and SHARK, respectively. As suggested by Patton (2011), we consider the MSE and QLIKE as loss measures.

Electronic copy available at: https://ssrn.com/abstract=3089929

These two measures are indeed robust to measurement errors. We also report, on the last two lines, the average MSE and QLIKE computed for the 18 NYSE assets.

We note that, for $M = 39$, the coefficient β_1 estimated by the HARK is 0.5420 ± 0.0153, while that of the $\text{HAR}_{\log}$ is 0.4167 ± 0.0158, meaning that the former is correcting for the bias induced by estimation errors. As M increases, the bias reduces and we observe that the two estimates become closer. The same effect is observed on loss measures, with the relative difference between the two models being larger for $M = 39$, and then reducing as M increases. We also note that the signal-to-noise ratio estimated by both the HARK and the SHARK increases as M increases, since measurement errors become less relevant.

Figure (10), (11), (12), (13) show, for $M = 39, 78, 390$, filtered estimates of SHAR and SHARK time-varying parameters in the whole S&P500 sample. We do not report the $\beta_{2,t}$ coefficient, as the estimated parameters a_3, c_3 are close to zero for both the SHAR and the SHARK. We observe that the SHARK provides more responsive estimates, especially for $M = 39$. As M increases, the parameters filtered by the two models tend to show similar dynamics. Indeed, while the SHARK is able to filter out measurement errors, SHAR estimates are bias-attenuated by the noise. In Figure (13), we see that both models capture the residual heteroskedasticity that we observed in Figure (6). However, the time-varying variance estimated by the SHAR is larger than that of the SHARK. This is another consequence of measurement errors. Indeed, the SHAR does not disentangle the measurement error variance from that of model innovations, and provides filtered variance estimates that include the contribution of both terms.

The coefficient β_{1Q} of the HARQ is negative, in agreement with the empirical results of Bollerslev et al. (2016a). Since the correction term in the HARQ accounts for measurement errors, as M increases, β_{1Q} should become smaller in absolute value. In contrast, for $M = 390$, it is still significantly different form zero and it is close to the one computed for $M = 39$. This empirical finding is in agreement with what we found in the simulation study in the online appendix, and indicates that the HARQ term can also capture nonlinear dependencies.

Further insights related to the effect of measurement errors are given by the in-sample log-likelihoods. As expected, the SHARK has the largest estimated log-likelihood for all M's, implying better in-sample fit resulting from modeling both measurement errors and time-varying parameters. The estimated log-likelihood of the HARK is larger than that of the SHAR for $M = 39$ and $M = 78$, where measurement errors play a more relevant role.

3.2 Out-of-sample analysis

As underlined, the period used for the out-of-sample analysis on S&P500 data coincides with the last 2259 days, from December 2002 to June 2013. We divide this long period into two subsamples, the first spanning from December 2002 to August 2008, and the second going from September 2008 to June 2013. The latter

Electronic copy available at: https://ssrn.com/abstract=3089929

includes the 2008 financial crisis and is characterized by a larger volatility. We report average loss measures computed in these two subsamples. On NYSE stock data, the out-of-sample analysis is instead performed in the period between January and December 2014.

In our framework, the relative average loss between two models can be interpreted as the forecast gain resulting from taking into account different forms of misspecification of the HAR. Given a generic loss measure $L(\cdot)$, we define:

$$\varphi_{\text{het}} = \frac{\text{E}[L(\text{HAR}_{\text{log}})]}{\text{E}[L(\text{HAR})]} \tag{41}$$

$$\varphi^{(1)}_{\text{tvp}} = \frac{\text{E}[L(\text{SHAR})]}{\text{E}[L(\text{HAR}_{\text{log}})]}, \quad \varphi^{(2)}_{\text{tvp}} = \frac{\text{E}[L(\text{SHARK})]}{\text{E}[L(\text{HARK})]} \tag{42}$$

$$\varphi^{(1)}_{\text{err}} = \frac{\text{E}[L(\text{HARK})]}{\text{E}[L(\text{HAR}_{\text{log}})]}, \quad \varphi^{(2)}_{\text{err}} = \frac{\text{E}[L(\text{SHARK})]}{\text{E}[L(\text{SHAR})]} \tag{43}$$

The first measure, φ_{het}, quantifies the forecast gain resulting from removing part of the heteroskedasticity of the HAR residuals. The other two measures, $\varphi^{(1)}_{\text{tvp}}$ and $\varphi^{(2)}_{\text{tvp}}$, quantify the effect of modeling the residual nonlinearities by introducing time-varying parameters. Finally, $\varphi^{(1)}_{\text{err}}$ and $\varphi^{(2)}_{\text{err}}$ assess the impact of measurement errors. Table (4) shows the results of the out-of-sample analysis on S&P500 data, while table (5) shows the results on NYSE stock data. The forecast gains defined in eq. (41)-(43) are instead reported in tables (7), (8).

We first discuss the results obtained on S&P500 data. In the first subsample, from December 2002 to August 2008, the SHARK provides the lowest MSE and QLIKE. The only exception is for $M = 390$. In this case, the SHAR provides a slightly lower MSE, since measurement errors are less relevant. Except for $M = 390$, where the HARQ has slightly lower MSE loss, the HARK always outperforms the latter in terms of both MSE and QLIKE. Compared to the basic HAR_{log} model, all the three extensions considered here provide lower MSE and QLIKE. We also note that the HARST improves over the HAR_{log}, but is outperformed by the SHAR.

If one uses the QLIKE as a loss measure, the three HAR extensions and the HARST are the only models belonging to the 90% model confidence set (Hansen et al. 2011), that we denote by $\widehat{\mathcal{M}}_{90\%}$. If one uses the MSE, the HARK and the SHARK are the only models included in $\widehat{\mathcal{M}}_{90\%}$ for $M = 39, 78$. For $M = 390$, all the models but the ARFIMA are included if a rolling window is employed. If a static window is used, only the SHAR and the SHARK are included. Finally, in most of the cases, the HAR_{log} has lower MSE and QLIKE compared to the HARQ.

In the second subsample, characterized by a larger volatility, the HARK and the SHARK often provide the lowest loss measures. In particular, for $M = 39$, the HARK is the only model included in $\widehat{\mathcal{M}}_{90\%}$. For $M = 78$, the SHARK is also included if one considers the QLIKE as a loss measure. For $M = 390$, other

Electronic copy available at: https://ssrn.com/abstract=3089929

nonlinear specifications like the SHAR and the HARST are also included. The volatility peaks observed in this subsample are associated to a large uncertainty. The HARK and the SHARK are successful in underrating volatility forecasts when variance estimates are affected by large measurement errors. This mechanism explains the nice performance of the two models in this subsample.

On stock data, the HARK, SHAR and SHARK improve significantly over the $\text{HAR}_{\log}$, as indicated by the two loss measures, and by the number of times each model is included in $\widehat{\mathcal{M}}_{90\%}$. With only two exceptions, the SHARK is always included in $\widehat{\mathcal{M}}_{90\%}$, while the HARK and the SHAR are included more frequently than the other benchmark models.

In the second subsample of &P500 data, and on NYSE stock data, the relative QLIKE of the HARQ is larger than one. The performance of the HAR and the $\text{HAR}^{\Delta J^{2\pm}}$ on stock data also deteriorates substantially compared to the $\text{HAR}_{\log}$. The reason is due to the estimation window including the 2008 financial crisis, which strongly affects the OLS estimates of the two models. The $\text{HAR}_{\log}$ and all the other models estimated on $\log(RV_t)$ are instead less affected by these huge peaks.

The SHAR and the SHARK have built-in time-varying HAR parameters that adapt automatically to changes in the persistence of volatility. As a consequence, when they are estimated on a static window, the relative performance with respect to models with constant HAR parameters estimated on a rolling window does not change substantially. Table 6 shows out-of-sample MSE and QLIKE obtained in the first subsample of S&P500 data by estimating the SHAR and the SHARK on a static window, and the remaining models on a rolling window. Compared to table (4), we see that the ranking of the models is unchanged. Similar results are obtained in the second subsample and on NYSE stock data.

From tables (7), (8), we note that the two forecast gains $\varphi_{\text{err}}^{(1)}$ and $\varphi_{\text{err}}^{(2)}$ tend to approach one as M increases, as one would expect from a model which exclusively accounts for measurement errors. On the first subsample of S&P500 data, $\varphi_{\text{err}}^{(1)}$ and $\varphi_{\text{err}}^{(2)}$ are on average smaller for $M = 78$ than for $M = 39$. This result suggests that the estimator of the noise variance in eq. (7) can provide less precise estimates of the true variance when computed al lower frequencies. The forecast gains φ_{het} resulting from taking into account heteroskedasticity are substantial, especially in the second subsample of S&P500 data, characterized by a larger volatility. On NYSE stock data, these forecast gains are slightly larger. Forecast gains resulting from modeling the residual nonlinear dependencies through time-varying parameters are roughly equal to 2% on average, as can be seen by looking at the values of $\varphi_{\text{tvp}}^{(1)}$ and $\varphi_{\text{tvp}}^{(2)}$ in tables (7), (8).

Our empirical findings reveal substantial improvements resulting from correcting for the main types of misspecification of the HAR. In particular, heteroskedasticity has deep impacts on HAR volatility forecasts, especially when the model is estimated on turbulent periods. Measurement errors are relevant at small (1-minute) and intermediate (5-minute) sampling frequencies, while modeling time-varying parameters results in forecast gains which are roughly equal to 2%.

Electronic copy available at: https://ssrn.com/abstract=3089929

We conclude this section by examining in more detail what kind of misspecification is captured by the HARQ. We found, both on simulated and empirical data, that the HARQ can capture measurement errors and nonlinear dependencies. In order to quantify the extent to which this happens, we regress the out-of-sample forecast errors of the HARQ against the forecast errors provided by all other models. Table (9) shows the R^2 obtained from such regression on the first subsample of S&P500 data, for $M = 39, 78, 390$. The forecast errors of the HARQ are better described in terms of the forecast errors of the SHARK, for $M = 39, 78$. At these sampling frequencies, the HARQ captures both measurement errors and nonlinear dependencies. This is also confirmed by the fact that the R^2 resulting from regressing against the HARK and the SHAR is larger than the one resulting from regressing against the $\text{HAR}_{\log}$, which neglects measurement errors and partly accounts for nonlinear dependencies. For $M = 390$, the largest R^2 is the one obtained by regressing against the SHAR, while the relative difference between regressions onto the $\text{HAR}_{\log}$ and the HARK forecast errors is lower. This result is not surprising, since at large sampling frequencies measurement errors become less important and the forecast gains provided by the HARQ are entirely due to capturing nonlinear dependencies.

3.3 Longer forecast horizons

In this Section, we assess the effect of measurement errors and nonlinear dependencies on out-of-sample volatility forecasts for longer time horizons. In particular, we choose a weekly ($j = 5$) and a monthly ($j = 22$) forecast horizon.

Bias-corrected, multi-step-ahead forecasts of the HARK, SHAR and SHARK models can be evaluated as described in Section (2.7). The multi-step-ahead forecasts of the HARST are computed by slightly modifying the formula of the $\text{HAR}_{\log}$ to also include the nonlinear lags. In absence of a dynamic specification for the quarticity term in eq. (9), longer horizon forecasts for the HARQ model can only be evaluated directly, i.e. by replacing the daily RV_t on the left-hand-side of eq. (9) with the variance aggregated at different frequencies. Correspondingly, a correction term is introduced to adjust the lag at the specific forecast horizon, as described by Bollerslev et al. 2016a. A similar approach is used to obtain multi-step-ahead forecasts of the $\text{HAR}^{\Delta J^{2\pm}}$ model.

Table (10) shows the relative out-of-sample MSE and QLIKE obtained on S&P500 data. We compute the loss measures in the period between December 2002 and June 2013, based both on a static and a rolling window. Table (11) reports the forecast gains defined in eq. (41)-(43). The SHARK provides lowest MSE and QLIKE in most of the cases, and it is always included in $\widehat{\mathcal{M}}_{90\%}$. The ARFIMA, being a long memory process, performs particularly well at long horizons, but is often outperformed by the SHARK. With only one exception, the HARK and the SHAR outperform the $\text{HAR}_{\log}$ and are included in $\widehat{\mathcal{M}}_{90\%}$ in most of the cases. The HARQ improves over the HAR, but is outperformed by the $\text{HAR}_{\log}$, which provides better

Electronic copy available at: https://ssrn.com/abstract=3089929

forecasts, especially on the monthly horizon.

Measurement errors have a similar effect as in daily forecasts, with $\varphi^{(1)}_{\text{err}}$ and $\varphi^{(2)}_{\text{err}}$ approaching one as M increases. Note, however, that the forecast gains increase as j increases, since longer horizon forecasts benefit from the higher persistence estimated by both the HARK and SHARK. A similar dependence on the forecast horizon is observed when modeling nonlinear dependencies. Figure (14) shows MSE-based forecast gains φ_{het}, $\varphi^{(1)}_{\text{err}}$, $\varphi^{(1)}_{\text{tvp}}$, for $j = 2, 4, \ldots, 22$ and $M = 78$. Both $\varphi^{(1)}_{\text{err}}$ and $\varphi^{(1)}_{\text{tvp}}$ exhibit a weak dependence on j, with the two measures declining slowly as the forecast horizon increases. In contrast, φ_{het} has a strong declining pattern that is due to the accumulation of forecast errors over large horizons.

The statistical significance of these forecasting gains is jointly tested through a pairwise version of the multi-horizon test recently proposed by Quaedvlieg (2019). In particular, as shown in table (12), we find strong evidence of uniform superior predictive ability for φ_{het}, by testing HAR$-$HAR$_{\log}$ MSE losses, and for $\varphi^{(1)}_{\text{err}}$, by testing HAR$_{\log}-$HARK MSE losses. In these two cases, the test achieves basically zero p-values. For φ_{tvp}, the test rejects the null hypothesis at the 5% c.l. when testing HAR$_{\log}-$SHAR MSE losses.

3.4 Robustness checks

The discussion in Section (2.5) showed that, even on $\log RV_t$ series, the HAR is misspecified, since its residuals are weakly heteroskedastic and parameters are time-varying. The HARK corrects HAR$_{\log}$ forecasts for measurement errors through the Kalman filter. However, being the latter based on a linear state-space representation, it is useful to quantify the misspecification of the HARK due to the above-mentioned effects. The forecast gain $\varphi^{(2)}_{\text{tvp}}$ in Eq. (42) is a possible indicator, since it quantifies the gain resulting from modeling heteroskedasticity and time-varying parameters in the HARK.

Another possible method is to observe that the assumption $h_t = V_t$ in Section (2.5) is fair only if the HARK is correctly specified. Thus, one can use a more flexible specification for h_t and check, on real data, the extent to which the latter deviates from the assumption based on correct specification. In particular, we have examined the following dynamic specifications for h_t:

1. $h_t = \alpha V_t + (1-\alpha)V_{t-1}$
2. $h_t = \alpha V_t + (1-\alpha)\text{mean}(V_t)$
3. $h_t = \alpha V_t + (1-\alpha)\text{median}(V_t)$
4. $h_t = \beta V_t$

where $0 \leq \alpha \leq 1$, $\beta > 0$ and mean$(\cdot)$ and median$(\cdot)$ denote the sample mean and median, respectively. In the first case, h_t is exponentially smoothed with past observations of V_t, and the parameter α indicates

Electronic copy available at: https://ssrn.com/abstract=3089929

the level of smoothing. In the second and third case, if $\alpha < 1$, h_t is shrunk towards its mean or median level. These two specifications can also be viewed as heteroskedasticity tests for h_t, in the sense that, if $\alpha > 0$, we reject the null assumption that h_t is homoskedastic, at a given confidence level. Finally, in the last case, β is a scaling constant that fine-tunes h_t in case the model is misspecified.

We estimated the HARK and the SHARK, with the dynamic specifications $\{1, 2, 3, 4\}$ for h_t, on the S&P500 sample. For each specification involving the parameter α, the latter turns out to be very close to one. For instance, for the specification 2, we found $\alpha = 0.999994$, and similar values have been recovered for the other specifications. This indicates that today's V_t is the best proxy for h_t and that there are not significant advantages in using lagged V_t's. Also, the null assumption of homoskedasticity for h_t is rejected, and this is in accordance with with what Bollerslev et al. (2016a) found on RV_t series.

Table (13) shows maximum likelihood estimates of the parameter β, for $M = 39, 78, 390$. We note that the estimated β is larger than one for the HARK, indicating significant departures from the assumption of correct specification. In particular, the fact that the estimated signal-to-noise ratio turns out to be lower, is due to deviations from the linearity assumption, which are spuriously interpreted by the Kalman filter as an excess of noise. It is worth underling that these more flexible specifications do not lead to significant out-of-sample improvements compared to the standard HARK model.

In the case of the SHARK, β is close to one for all the three sampling frequencies, in agreement with the assumption of correct specification. This result is in accordance with what we found in Section (2.6) when looking at the normalized residuals of the SHAR: implementing score-driven parameters in the $\text{HAR}_{\log}$ results in removing the misspecification due to heteroskedasticity. The SHARK, which in turn corrects SHAR forecasts for measurement errors, provides the most complete specification for realized variance.

4 Conclusions

In this paper, we proposed new HAR extensions accounting for measurement errors and nonlinear dependencies in volatility dynamics. In particular, by considering the $\text{HAR}_{\log}$ as a starting point, we devised three models that progressively capture the two effects. The HARK adjusts $\text{HAR}_{\log}$ parameters for the attenuation bias induced by measurement errors. This is done by correcting RV_t forecasts through a time-varying Kalman gain driven by realized quarticity. The SHAR has time-varying parameters and features i.i.d. standardized residuals. Finally, the SHARK is a generalization of the HARK that allows for time-varying parameters. It provides more responsive parameter dynamics, as they are not affected by measurement errors. The relative improvement of a model compared to another can thus be used to quantify the importance of removing each form of misspecification.

We provided simulation and empirical evidence that the effects captured by these new models are

Electronic copy available at: https://ssrn.com/abstract=3089929

statistically significant, since they translate into superior out-of-sample forecasts compared to the basic HAR_{log} specification and other competing approaches. For instance, on real data, the analysis based on the model confidence set includes our new extensions on the confidence set and excludes the remaining models in the vast majority of the cases. Interestingly, the performance of the SHAR and the SHARK does not change significantly when they are estimated on a static window or on a rolling window, since they have built-in time-varying HAR coefficients which automatically adapt to changes in persistence.

As a final outcome of our analysis, we conclude that measurement errors are relevant at small and intermediate sampling frequencies. Their corresponding forecast gains slightly increase with the forecast horizon. Accounting for time-varying parameters provides statistically significant forecast gains. The latter are independent from the sampling frequency, and slightly increase with the forecast horizon.

Acknowledgements

We are particularly grateful for suggestions we received from Francisco Blasques, Tim Bollerslev, Andrew J. Patton, Rogier Quaedvlieg, Roberto Renò and seminar participants at Vrije Universiteit Amsterdam and the 11th CFE conference in London.

References

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

Andersen, T. G., Bollerslev, T., Diebold, F. X., 2007. Roughing it up: Including jump components in the measurement, modeling, and forecasting of return volatility. The Review of Economics and Statistics 89 (4), 701–720.

Andersen, T. G., Bollerslev, T., Diebold, F. X., Labys, P., 2003. Modeling and forecasting realized volatility. Econometrica 71 (2), 579–625.

Asai, M., McAleer, M., Medeiros, M. C., 2012. Modelling and forecasting noisy realized volatility. Computational Statistics & Data Analysis 56 (1), 217 – 230.

Barndorff-Nielsen, O. E., Shephard, N., 2002. Econometric analysis of realized volatility and its use in estimating stochastic volatility models. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 64 (2), 253–280.

Bekierman, J., Manner, H., 2018. Forecasting realized variance measures using time-varying coefficient models. International Journal of Forecasting 34 (2), 276 – 287.

Electronic copy available at: https://ssrn.com/abstract=3089929

Blasques, F., Koopman, S. J., Lucas, A., Aug. 2014. Optimal formulations for nonlinear autoregressive processes. Working paper Available at https://ssrn.com/abstract=2478575.

Bollerslev, T., Patton, A. J., Quaedvlieg, R., 2016a. Exploiting the errors: A simple approach for improved volatility forecasting. Journal of Econometrics 192 (1), 1 – 18.

Bollerslev, T., Patton, A. J., Quaedvlieg, R., Apr. 2016b. Modeling and Forecasting (Un)Reliable Realized Covariances for More Reliable Financial Decisions. CREATES Research Papers 2016-10, Department of Economics and Business Economics, Aarhus University.

Chen, X. B., Gao, J., Li, D., Silvapulle, P., 2018. Nonparametric estimation and forecasting for time-varying coefficient realized volatility models. Journal of Business & Economic Statistics 36 (1), 88–100.

Chen, Y., Härdle, W. K., Pigorsch, U., 2010. Localized realized volatility modeling. Journal of the American Statistical Association 105 (492), 1376–1393.

Cipollini, F., Gallo, G., Otranto, E., Sep. 2017. On heteroskedasticity and regimes in volatility forecasting. Working paper Available at https://ssrn.com/abstract=3037550.

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

Corsi, F., Fusari, N., Vecchia, D. L., 2013. Realizing smiles: Options pricing with realized volatility. Journal of Financial Economics 107 (2), 284 – 304.

Corsi, F., Mittnik, S., Pigorsch, C., Pigorsch, U., 2008. The volatility of realized volatility. Econometric Reviews 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.

Creal, D., Koopman, S. J., Lucas, A., 2008. A general framework for observation driven time-varying parameter models. SSRN Electronic Journal.

Creal, D., Koopman, S. J., Lucas, A., 2013. Generalized autoregressive score models with applications. Journal of Applied Econometrics 28 (5), 777–795.

Creal, D., Schwaab, B., Koopman, S. J., Lucas, A., 2014. Observation-driven mixed-measurement dynamic factor models with an application to credit risk. The Review of Economics and Statistics 96 (5), 898–915.

Delle Monache, D., Petrella, I., Venditti, F., 2016. Adaptive state space models with applications to the business cycle and financial stress. CEPR Discussion Paper (DP11599).

Electronic copy available at: https://ssrn.com/abstract=3089929

Durbin, J., Koopman, S., 2012. Time Series Analysis by State Space Methods: Second Edition. Oxford Statistical Science Series. OUP Oxford.

Engle, R., 2002. New frontiers for arch models. Journal of Applied Econometrics 17 (5), 425–446.

Engle, R. F., Gallo, G. M., 2006. A multiple indicators model for volatility using intra-daily data. Journal of Econometrics 131 (1), 3 – 27.

Hansen, P. R., Lunde, A., Nason, J. M., 2011. The model confidence set. Econometrica 79 (2), 453–497.

Harvey, A., 1991. Forecasting, Structural Time Series Models and the Kalman Filter. Cambridge University Press.

Harvey, A. C., 2013. Dynamic Models for Volatility and Heavy Tails: With Applications to Financial and Economic Time Series. Econometric Society Monographs. Cambridge University Press.

McAleer, M., Medeiros, M. C., 2008. A multiple regime smooth transition heterogeneous autoregressive model for long memory and asymmetries. Journal of Econometrics 147 (1), 104 – 119.

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

Patton, A. J., Sheppard, K., 2015. Good volatility, bad volatility: Signed jumps and the persistence of volatility. The Review of Economics and Statistics 97 (3), 683–697.

Protter, P., 1992. Stochastic Integration and Differential Equation, 2nd Edition. Springer-Verlag, Berlin, Heidelberg.

Quaedvlieg, R., 2019. Multi-horizon forecast comparison. Journal of Business & Economic Statistics, Forthcoming.

Simon, D., 2006. Kalman filter generalizations. John Wiley & Sons, Inc., pp. 183–227.

Wansbeek, T., Meijer, E., 2000. Measurement Error and Latent Variables in Econometrics, 1st Edition. Elsevier.

Electronic copy available at: https://ssrn.com/abstract=3089929

				$M = 39$					
	HAR	HARQ	$\text{HAR}_{\log}$	HARK	SHAR	SHARK	ARFIMA	HARST	$\text{HAR}^{\Delta J^{2\pm}}$
β_0	0.0879	-0.0402	-0.0372	-0.0309				-0.0421	0.0588
	(0.0264)	(0.0270)	(0.0119)	(0.0153)				(0.0191)	(0.0881)
β_1	0.3460	0.7068	0.4167	0.5420				0.4281	
	(0.0160)	(0.0284)	(0.0158)	(0.0338)				(0.0201)	
β_2	0.3962	0.3110	0.3655	0.2944				0.3830	0.4695
	(0.0229)	(0.0229)	(0.0219)	(0.0445)				(0.0287)	(0.2238)
β_3	0.1656	0.0947	0.1602	0.1165				0.1399	0.1602
	(0.0229)	(0.0227)	(0.0203)	(0.0308)				(0.0198)	(0.0973)
β_{1Q}		-0.0119							
		(0.0007)							
a_1, c_1					2.2996	3.6931			
					(0.0794)	(0.2418)			
a_2, c_2					0.2300	0.5442			
					(0.0804)	(0.0525)			
a_3, c_3					0.0000	0.0001			
					(0.0000)	(0.0002)			
a_4, c_4					0.0001	0.1944			
					(0.0003)	(0.0847)			
a_5, c_5					1.5477	1.9583			
					(0.1651)	(0.0415)			
Signal-to-Noise				3.0487		3.7064			
ψ							0.9630		
							(0.0125)		
χ							-0.8922		
							(0.0298)		
d							0.3361		
							(0.0345)		
N. of regimes								2	
β^{BV}									0.0354
									(0.0055)
β^{J^+}									0.4895
									(0.2128)
β^{J^-}									-1.3866
									(0.2591)
MSE	2.2490	2.0464	2.2108	2.2037	2.1689	2.1617	2.2120	2.2333	2.0190
QLIKE	0.1844	0.6328	0.1847	0.1837	0.1822	0.1783	0.1880	0.1894	0.1974
$-\text{LogL} \cdot 10^{-3}$				1.8585	1.8716	1.8455	2.3658		
$\overline{\text{MSE}}$	5.7066	5.5769	6.3001	6.2670	6.1582	6.2732	6.2350	6.2869	5.2111
$\overline{\text{QLIKE}}$	0.1772	0.1774	0.1732	0.1756	0.1739	0.1758	0.1748	0.1743	0.1769

Table 1: OLS estimates of HAR, HARQ, $\text{HAR}_{\log}$, $\text{HAR}^{\Delta J^{2\pm}}$ coefficients and maximum likelihood estimates of HARK, SHAR, SHARK, ARFIMA, HARST parameters obtained on S&P500 data computed with 10-minute returns. Robust standard errors are indicated inside parenthesis. We also show the in-sample MSE, QLIKE on S&P500 data and the average in-sample MSE, QLIKE on NYSE stock data. The SHAR and SHARK parameters are multiplied by 100.

Electronic copy available at: https://ssrn.com/abstract=3089929

				$M = 78$					
	HAR	HARQ	$\text{HAR}_{\log}$	HARK	SHAR	SHARK	ARFIMA	HARST	$\text{HAR}^{\Delta J^{2\pm}}$
β_0	0.0921	0.0127	-0.0307	-0.0270				-0.0367	0.0679
	(0.0280)	(0.0279)	(0.0106)	(0.0147)				(0.0097)	(0.0201)
β_1	0.3249	0.5780	0.4500	0.5480				0.4637	
	(0.0159)	(0.0237)	(0.0157)	(0.0299)				(0.0311)	
β_2	0.4166	0.3211	0.3531	0.2914				0.3724	0.3839
	(0.0232)	(0.0236)	(0.0214)	(0.0367)				(0.0125)	(0.0265)
β_3	0.1661	0.1309	0.1442	0.1147				0.1183	0.1498
	(0.0236)	(0.0232)	(0.0193)	(0.0276)				(0.0200)	(0.0189)
β_{1Q}		-0.0059							
		(0.0004)							
a_1, c_1					2.0385	3.3315			
					(0.0629)	(0.1300)			
a_2, c_2					0.3252	0.5295			
					(0.0358)	(0.0304)			
a_3, c_3					0.0000	0.0000			
					(0.0002)	(0.0013)			
a_4, c_4					0.0020	0.2346			
					(0.0021)	(0.0370)			
a_5, c_5					1.9973	1.7019			
					(0.0632)	(0.0368)			
Signal-to-Noise				5.0582		6.0415			
ψ							0.9566		
							(0.0158)		
χ							-0.8901		
							(0.0346)		
d							0.3758		
							(0.0377)		
N. of regimes								2	
β^{BV}									0.2684
									(0.0190)
β^{J^+}									0.2555
									(0.1871)
β^{J^-}									-1.0569
									(0.2300)
MSE	2.1577	1.9766	2.1549	2.1587	2.1431	2.1542	2.1683	2.1658	2.0418
QLIKE	0.1536	0.5941	0.1532	0.1507	0.1501	0.1465	0.1564	0.1579	0.1514
$-\text{LogL} \cdot 10^{-3}$				1.6737	1.6851	1.6561	1.8153		
$\overline{\text{MSE}}$	5.5758	5.3013	5.9656	5.8987	5.8485	5.9093	5.8658	5.9708	5.4268
$\overline{\text{QLIKE}}$	0.1452	0.1464	0.1400	0.1407	0.1401	0.1408	0.1412	0.1414	0.1458

Table 2: OLS estimates of HAR, HARQ, $\text{HAR}_{\log}$, $\text{HAR}^{\Delta J^{2\pm}}$ coefficients and maximum likelihood estimates of HARK, SHAR, SHARK, ARFIMA, HARST parameters obtained on S&P500 data computed with 5-minute returns. Robust standard errors are indicated inside parenthesis. We also show the in-sample MSE, QLIKE on S&P500 data and the average in-sample MSE, QLIKE on NYSE stock data. The SHAR and SHARK parameters are multiplied by 100.

Electronic copy available at: https://ssrn.com/abstract=3089929

				$M = 390$					
	HAR	HARQ	$\text{HAR}_{\log}$	HARK	SHAR	SHARK	ARFIMA	HARST	$\text{HAR}^{\Delta J^{2\pm}}$
β_0	0.0853	-0.0687	-0.0272	-0.0264				-0.0295	0.0632
	(0.0257)	(0.0246)	(0.0091)	(0.0138)				(0.0091)	(0.0188)
β_1	0.3539	0.8679	0.4996	0.5428				0.5046	
	(0.0160)	(0.0257)	(0.0155)	(0.0273)				(0.2238)	
β_2	0.4166	0.2652	0.3442	0.3161				0.3665	0.4690
	(0.0227)	(0.0219)	(0.0206)	(0.0343)				(0.0279)	(0.0322)
β_3	0.1429	0.0131	0.1056	0.0931				0.0865	0.1201
	(0.0220)	(0.0211)	(0.0176)	(0.0269)				(0.0111)	(0.0233)
β_{1Q}		-0.0104							
		(0.0004)							
a_1, c_1					2.7082	2.8887			
					(0.2107)	(0.0212)			
a_2, c_2					1.0119	0.9008			
					(0.1249)	(0.0139)			
a_3, c_3					0.0000	0.0000			
					(0.0028)	(0.0064)			
a_4, c_4					0.1531	0.3086			
					(0.0298)	(0.0093)			
a_5, c_5					2.1238	1.7671			
					(0.1465)	(0.0719)			
Signal-to-Noise				16.5013		19.0204			
ψ							0.9442		
							(0.0214)		
χ							-0.8706		
							(0.0446)		
d							0.4102		
							0.0438		
N. of regimes								2	
β^{BV}									0.1802
									(0.0210)
β^{J^+}									0.6952
									(0.2063)
β^{J^-}									-2.6893
									(0.5010)
MSE	2.3724	1.9556	2.3027	2.3058	2.3188	2.2867	2.2887	2.3496	2.2198
QLIKE	0.1246	0.3402	0.1210	0.1194	0.1191	0.1174	0.1222	0.1244	0.1264
$-\text{LogL} \cdot 10^{-3}$				1.3861	1.3674	1.3540	1.5444		
$\overline{\text{MSE}}$	4.0100	3.8106	4.1341	4.1382	4.1397	4.0882	4.0984	4.1923	3.9311
$\overline{\text{QLIKE}}$	0.1011	0.0997	0.0962	0.0966	0.0963	0.0960	0.0988	0.0985	0.0988

Table 3: OLS estimates of HAR, HARQ, $\text{HAR}_{\log}$, $\text{HAR}^{\Delta J^{2\pm}}$ coefficients and maximum likelihood estimates of HARK, SHAR, SHARK, ARFIMA, HARST parameters obtained on S&P500 data computed with 1-minute returns. Robust standard errors are indicated inside parenthesis. We also show the in-sample MSE, QLIKE on S&P500 data and the average in-sample MSE, QLIKE on NYSE stock data. The SHAR and SHARK parameters are multiplied by 100.

Electronic copy available at: https://ssrn.com/abstract=3089929

		HAR	HARQ	HAR_{log}	HARK	SHAR	SHARK	ARFIMA	HARST	$\text{HAR}^{\Delta J^{2\pm}}$
		Dec. 2002 - Aug. 2008								
		$\underline{M = 39}$								
MSE	RW	1.0000	0.9897	0.9927	0.9699*	0.9758	**0.9488***	0.9970	0.9868	0.9799
	SW	1.0000	0.9823	0.9892	0.9668*	0.9712	**0.9462***	0.9925	0.9821	0.9702
QLIKE	RW	1.0000	0.9037	0.8863	0.8765	0.8628*	**0.8503***	0.8810	0.8798	0.9335
	SW	1.0000	0.9043	0.8840	0.8701	0.8587*	**0.8483***	0.8759	0.8629*	0.9312
		$\underline{M = 78}$								
MSE	RW	1.0000	0.9714	0.9555	0.9284*	0.9430	**0.9142***	0.9654	0.9555	0.9747
	SW	1.0000	0.9685	0.9534	0.9232*	0.9311	**0.9025***	0.9562	0.9533	0.9711
QLIKE	RW	1.0000	0.8689	0.7844	0.7668*	0.7706*	**0.7421***	0.7792	0.7844	0.8996
	SW	1.0000	0.8653	0.7823	0.7673*	0.7638*	**0.7362***	0.7783	0.7820	0.8950
		$\underline{M = 390}$								
MSE	RW	1.0000*	0.9595*	0.9781*	0.9636*	**0.9502***	0.9527*	0.9964	0.9712*	0.9542*
	SW	1.0000	0.9567	0.9770	0.9625	**0.9440***	0.9452*	0.9981	0.9677	0.9556
QLIKE	RW	1.0000	0.9244	0.9022	0.8911	0.8795*	**0.8725***	0.8991	0.8867*	0.8934
	SW	1.0000	0.9240	0.9032	0.8897	0.8783*	**0.8711***	0.8923	0.8832*	0.8922
		Sep. 2008 - Jun. 2013								
		$\underline{M = 39}$								
MSE	RW	1.0000	0.9814	0.5778	**0.5293***	0.5663	0.5619	0.5899	0.5732	1.0112
	SW	1.0000	0.9688	0.5693	**0.5173***	0.5610	0.5546	0.5761	0.5655	1.0855
QLIKE	RW	1.0000	1.1298	0.9438	**0.9294***	0.9439	0.9403	0.9414	0.9411	1.0097
	SW	1.0000	1.1270	0.9451	**0.9244***	0.9405	0.9359	0.9386	0.9413	1.0115
		$\underline{M = 78}$								
MSE	RW	1.0000	0.9428	0.6631	**0.6300***	0.6661	0.6601	0.6720	0.6621	1.0565
	SW	1.0000	0.9388	0.6630	**0.6287***	0.6604	0.6551	0.6680	0.6623	1.0495
QLIKE	RW	1.0000	1.1087	0.9395	**0.9230***	0.9381	0.9307*	0.9386	0.9346	0.9912
	SW	1.0000	1.1032	0.9310	**0.9162***	0.9390	0.9231*	0.9319	0.9301	0.9905
		$\underline{M = 390}$								
MSE	RW	1.0000	0.9757	0.7256	**0.7051***	0.7112*	0.7117*	0.7371	0.7216	1.0133
	SW	1.0000	0.9680	0.7144	**0.7028***	0.7067*	0.7081*	0.7187	0.7121*	1.0082
QLIKE	RW	1.0000	1.1103	0.9166	0.9097*	0.9120*	**0.9077***	0.9238	0.9141	1.0195
	SW	1.0000	1.1101	0.9078	0.9010*	0.9002*	**0.8987***	0.9210	0.9061	1.0158

Table 4: Relative out-of-sample losses of HAR, HARQ, HAR_{log}, HARK, SHAR, SHARK, ARFIMA, HARST, $\text{HAR}^{\Delta J^{2\pm}}$ on S&P500 data in the two subsamples [Dec. 2002 - Aug. 2008] and [Sep. 2008 - Jun. 2013]. We report losses computed both on a rolling window (RW) and on a static window (SW). The presence of an asterisk indicates that the model is included in $\widehat{\mathcal{M}}_{90\%}$.

Electronic copy available at: https://ssrn.com/abstract=3089929

		HAR	HARQ	HAR_{log}	HARK	SHAR	SHARK	ARFIMA	HARST	$\text{HAR}^{\Delta J^{2\pm}}$
						$\underline{M = 39}$				
MSE	RW	$1.0000^{(3)}$	$0.7651^{(5)}$	$0.5639^{(5)}$	$0.5433^{(14)}$	$0.5488^{(14)}$	$\mathbf{0.5351}^{(18)}$	$0.5548^{(5)}$	$0.5629^{(5)}$	$0.9182^{(3)}$
	SW	$1.0000^{(2)}$	$0.7601^{(6)}$	$0.5529^{(5)}$	$0.5385^{(14)}$	$0.5404^{(15)}$	$\mathbf{0.5278}^{(18)}$	$0.5534^{(5)}$	$0.5557^{(5)}$	$0.9102^{(4)}$
QLIKE	RW	$1.0000^{(3)}$	$0.9579^{(3)}$	$0.7366^{(10)}$	$0.7295^{(14)}$	$0.7221^{(16)}$	$\mathbf{0.7167}^{(17)}$	$0.7310^{(11)}$	$0.7370^{(8)}$	$0.9622^{(3)}$
	SW	$1.0000^{(3)}$	$0.9521^{(3)}$	$0.7299^{(11)}$	$0.7207^{(15)}$	$0.7179^{(16)}$	$\mathbf{0.7125}^{(18)}$	$0.7288^{(11)}$	$0.7351^{(8)}$	$0.9586^{(3)}$
						$\underline{M = 78}$				
MSE	RW	$1.0000^{(2)}$	$0.9034^{(3)}$	$0.6612^{(7)}$	$0.6485^{(12)}$	$0.6487^{(15)}$	$\mathbf{0.6395}^{(18)}$	$0.6631^{(7)}$	$0.6647^{(6)}$	$0.9793^{(2)}$
	SW	$1.0000^{(2)}$	$0.9021^{(3)}$	$0.6601^{(7)}$	$0.6427^{(12)}$	$0.6477^{(15)}$	$\mathbf{0.6380}^{(18)}$	$0.6658^{(6)}$	$0.6587^{(7)}$	$0.9685^{(2)}$
QLIKE	RW	$1.0000^{(3)}$	$1.2628^{(1)}$	$0.7071^{(11)}$	$0.7010^{(13)}$	$0.6932^{(17)}$	$\mathbf{0.6866}^{(18)}$	$0.7018^{(10)}$	$0.7091^{(10)}$	$1.2922^{(1)}$
	SW	$1.0000^{(2)}$	$1.2787^{(1)}$	$0.7015^{(12)}$	$0.6989^{(13)}$	$0.6912^{(17)}$	$\mathbf{0.6818}^{(18)}$	$0.7019^{(11)}$	$0.6988^{(13)}$	$1.2788^{(1)}$
						$\underline{M = 390}$				
MSE	RW	$1.0000^{(1)}$	$0.8492^{(1)}$	$0.5437^{(3)}$	$0.5691^{(3)}$	$0.5287^{(9)}$	$\mathbf{0.5193}^{(18)}$	$0.5457^{(3)}$	$0.5598^{(3)}$	$0.9957^{(1)}$
	SW	$1.0000^{(1)}$	$0.8326^{(2)}$	$0.5337^{(3)}$	$0.5525^{(3)}$	$0.5221^{(10)}$	$\mathbf{0.5109}^{(18)}$	$0.5401^{(2)}$	$0.5466^{(3)}$	$0.9866^{(1)}$
QLIKE	RW	$1.0000^{(2)}$	$2.3093^{(1)}$	$0.6002^{(10)}$	$0.5972^{(14)}$	$0.5920^{(17)}$	$\mathbf{0.5826}^{(17)}$	$0.6055^{(8)}$	$0.6020^{(10)}$	$1.0684^{(1)}$
	SW	$1.0000^{(2)}$	$2.3020^{(1)}$	$0.5921^{(10)}$	$0.5848^{(12)}$	$0.5901^{(18)}$	$\mathbf{0.5823}^{(18)}$	$0.6088^{(7)}$	$0.6001^{(9)}$	$1.0536^{(1)}$

Table 5: Relative out-of-sample losses of HAR, HARQ, HAR_{log}, HARK, SHAR, SHARK, ARFIMA, HARST, $\text{HAR}^{\Delta J^{2\pm}}$ averaged over the 18 NYSE stocks. We report average losses computed both on a rolling window (RW) and on a static window (SW). We show in parenthesis the number of times each model is included in $\widehat{\mathcal{M}}_{90\%}$.

	HAR	HARQ	HAR_{log}	HARK	SHAR	SHARK	ARFIMA	HARST	$\text{HAR}^{\Delta J^{2\pm}}$
	Dec. 2002 - Aug. 2008								
					$\underline{M = 39}$				
MSE	1.0000	0.9897	0.9927	$0.9699^{\star}$	0.9770	$\mathbf{0.9508}^{\star}$	0.9970	0.9868	0.9799
QLIKE	1.0000	0.9037	0.8863	0.8765	$0.8643^{\star}$	$\mathbf{0.8536}^{\star}$	0.8810	0.8798	0.9335
					$\underline{M = 78}$				
MSE	1.0000	0.9714	0.9555	$0.9284^{\star}$	0.9452	$\mathbf{0.9163}^{\star}$	0.9654	0.9555	0.9747
QLIKE	1.0000	0.8689	0.7844	$0.7668^{\star}$	$0.7733^{\star}$	$\mathbf{0.7455}^{\star}$	0.7792	0.7844	0.8996
					$\underline{M = 390}$				
MSE	$1.0000^{\star}$	$0.9595^{\star}$	$0.9781^{\star}$	$0.9636^{\star}$	$\mathbf{0.9524}^{\star}$	$0.9538^{\star}$	0.9964	$0.9712^{\star}$	$0.9542^{\star}$
QLIKE	1.0000	0.9244	0.9022	0.8911	$0.8830^{\star}$	$\mathbf{0.8759}^{\star}$	0.8991	$0.8867^{\star}$	0.8934

Table 6: Relative out-of-sample losses of HAR, HARQ, HAR_{log}, HARK, SHAR, SHARK, ARFIMA, HARST, $\text{HAR}^{\Delta J^{2\pm}}$ on S&P500 data in the subsample [Dec. 2002 - Aug. 2008]. The SHAR and the SHARK are estimated on a static window, while the remaining models are estimated on a rolling window. The presence of an asterisk indicates that the model is included in $\widehat{\mathcal{M}}_{90\%}$.

Electronic copy available at: https://ssrn.com/abstract=3089929

		φ_{het}	$\varphi^{(1)}_{\text{tvp}}$	$\varphi^{(2)}_{\text{tvp}}$	$\varphi^{(1)}_{\text{err}}$	$\varphi^{(2)}_{\text{err}}$
		Dec. 2002 - Aug. 2008				
				$\underline{M = 39}$		
MSE	RW	0.9927	0.9830	0.9782	0.9771	0.9723
	SW	0.9892	0.9818	0.9787	0.9774	0.9743
QLIKE	RW	0.8863	0.9735	0.9701	0.9890	0.9855
	SW	0.8840	0.9714	0.9749	0.9843	0.9879
				$\underline{M = 78}$		
MSE	RW	0.9555	0.9869	0.9847	0.9715	0.9695
	SW	0.9534	0.9766	0.9776	0.9683	0.9693
QLIKE	RW	0.7844	0.9824	0.9678	0.9775	0.9630
	SW	0.7823	0.9764	0.9595	0.9808	0.9639
				$\underline{M = 390}$		
MSE	RW	0.9781	0.9715	0.9887	0.9852	1.0026
	SW	0.9770	0.9662	0.9820	0.9852	1.0013
QLIKE	RW	0.9022	0.9748	0.9791	0.9876	0.9920
	SW	0.9032	0.9724	0.9791	0.9851	0.9918
		Sep. 2008 - Jun. 2013				
				$\underline{M = 39}$		
MSE	RW	0.5778	0.9801	1.0616	0.9161	0.9922
	SW	0.5693	0.9854	1.0721	0.9087	0.9886
QLIKE	RW	0.9438	1.0001	1.0117	0.9847	0.9962
	SW	0.9451	0.9951	1.0124	0.9781	0.9951
				$\underline{M = 78}$		
MSE	RW	0.6631	1.0045	1.0478	0.9501	0.9910
	SW	0.6630	0.9961	1.0420	0.9483	0.9920
QLIKE	RW	0.9395	0.9985	1.0083	0.9824	0.9921
	SW	0.9310	1.0086	1.0075	0.9841	0.9931
				$\underline{M = 390}$		
MSE	RW	0.7256	0.9802	1.0094	0.9717	1.0007
	SW	0.7144	0.9892	1.0075	0.9838	1.0020
QLIKE	RW	0.9166	0.9950	0.9978	0.9925	0.9953
	SW	0.9078	0.9916	0.9976	0.9925	0.9983

Table 7: Forecast gains φ_{het}, $\varphi^{(1)}_{\text{tvp}}$, $\varphi^{(2)}_{\text{tvp}}$, $\varphi^{(1)}_{\text{err}}$, $\varphi^{(2)}_{\text{err}}$ defined in eq. (41), (42), (43) computed on S&P500 data using the MSE and QLIKE as loss measures. We denote by RW the forecast gains computed on a rolling window and by SW those computed on a static window. Results for both subsamples [Dec. 2002 - Aug. 2008] and [Sep. 2008 - Jun. 2013] are reported.

Electronic copy available at: https://ssrn.com/abstract=3089929

		φ_{het}	$\varphi^{(1)}_{\text{tvp}}$	$\varphi^{(2)}_{\text{tvp}}$	$\varphi^{(1)}_{\text{err}}$	$\varphi^{(2)}_{\text{err}}$
				$\underline{M = 39}$		
MSE	RW	0.5639	0.9732	0.9849	0.9635	0.9750
	SW	0.5529	0.9776	0.9801	0.9740	0.9767
QLIKE	RW	0.7366	0.9803	0.9825	0.9904	0.9925
	SW	0.7299	0.9836	0.9886	0.9874	0.9925
				$\underline{M = 78}$		
MSE	RW	0.6612	0.9811	0.9861	0.9809	0.9858
	RW	0.6601	0.9812	0.9927	0.9736	0.9850
QLIKE	SW	0.7071	0.9803	0.9795	0.9914	0.9905
	SW	0.7015	0.9853	0.9755	0.9963	0.9864
				$\underline{M = 390}$		
MSE	RW	0.5437	0.9724	0.9125	1.0468	0.9822
	SW	0.5337	0.9783	0.9247	1.0352	0.9785
QLIKE	RW	0.6002	0.9863	0.9756	0.9951	0.9841
	SW	0.5921	0.9966	0.9957	0.9877	0.9868

Table 8: Forecast gains φ_{het}, $\varphi^{(1)}_{\text{tvp}}$, $\varphi^{(2)}_{\text{tvp}}$, $\varphi^{(1)}_{\text{err}}$, $\varphi^{(2)}_{\text{err}}$ defined in eq. (41), (42), (43) computed using the MSE and QLIKE as loss measures and averaged over the 18 NYSE stocks. We denote by RW the forecast gains computed on a rolling window and by SW those computed on a static window.

Electronic copy available at: https://ssrn.com/abstract=3089929

	HAR	HAR_{log}	HARK	SHAR	SHARK
			$\underline{M = 39}$		
HARQ	0.9490	0.9129	0.9281	0.9627	0.9741
			$\underline{M = 78}$		
HARQ	0.9321	0.9004	0.9235	0.9475	0.9737
			$\underline{M = 390}$		
HARQ	0.9395	0.9405	0.9517	0.9779	0.9623

Table 9: R^2 resulting from regressing the forecast errors of the HARQ against the forecast errors of all other models.

Electronic copy available at: https://ssrn.com/abstract=3089929

		HAR	HARQ	HAR_{log}	HARK	SHAR	SHARK	ARFIMA	HARST	$\text{HAR}^{\Delta J^{2\pm}}$
						j = 5				
						$M = 39$				
MSE	RW	1.0000	0.9519	0.9302	0.9047	0.9202	**0.8721***	0.8980	0.9680	0.9721
	SW	1.0000	0.9470	0.9280	0.8923	0.9125	**0.8622***	0.8922	0.9621	0.9688
QLIKE	RW	1.0000	0.9263	0.8253	0.8051	0.8031	**0.7713***	0.7980	0.8198	0.8462
	SW	1.0000	0.9201	0.8176	0.8001	0.7966	**0.7651***	0.7952	0.8147	0.8305
						$M = 78$				
MSE	RW	1.0000	0.9204	0.8853	0.8652	0.8601	**0.8442***	0.8542*	0.8853	0.9464
	SW	1.0000	0.9184	0.8798	0.8589*	0.8520*	**0.8399***	0.8507*	0.8795	0.9403
QLIKE	RW	1.0000	0.8917	0.7662	0.7452	0.7379*	**0.7222***	0.7406	0.7662	0.8058
	SW	1.0000	0.8899	0.7625	0.7421	0.7322*	**0.7178***	0.7382	0.7613	0.7926
						$M = 390$				
MSE	RW	1.0000	0.9422	0.9190	0.9084*	0.9128	**0.9031***	0.9038*	0.9221	0.9654
	SW	1.0000	0.9985	0.9126	0.9021*	0.9051	0.8953*	**0.8913***	0.9198	0.9555
QLIKE	RW	1.0000	0.9387	0.8572	0.8426	0.8339*	**0.8245***	0.8420	0.8644	0.8605
	SW	1.0000	0.9321	0.8530	0.8400	0.8275*	**0.8149***	0.8368	0.8591	0.8623
						j = 22				
						$M = 39$				
MSE	RW	1.0000	1.0081	0.8686	0.8334	0.8549	**0.7911***	0.8067*	0.8791	1.0023
	SW	1.0000	1.0111	0.8675	0.8288	0.8498	**0.7872***	0.8033	0.8710	1.0015
QLIKE	RW	1.0000	0.9755	0.8023	0.7630	0.8024	**0.6922***	0.7505	0.8061	0.8134
	SW	1.0000	0.9701	0.7997	0.7588	0.7956	**0.6853***	0.7459	0.7985	0.8122
						$M = 78$				
MSE	RW	1.0000	0.9896	0.8319	0.8039*	0.8050*	**0.7833***	0.7993*	0.8147	0.9688
	SW	1.0000	0.9822	0.8258	0.7952*	0.7975*	**0.7782***	0.7901*	0.8077	0.9631
QLIKE	RW	1.0000	0.9747	0.7697	0.7347*	0.7627	**0.6875***	0.7269*	0.7632	0.7848
	SW	1.0000	0.9690	0.7611	0.7388*	0.7545	**0.6809***	0.7123*	0.7598	0.7820
						$M = 390$				
MSE	RW	1.0000	1.0187	0.9409	0.9241*	0.9400	0.9000*	**0.8969***	0.9512	0.9942
	SW	1.0000	1.0222	0.9325	0.9248	0.9198	**0.8818***	0.8862*	0.9433	0.9887
QLIKE	RW	1.0000	1.0150	0.8712	0.8468	0.8441	**0.7981***	0.8539	0.9087	0.9008
	SW	1.0000	1.0171	0.8622	0.8432	0.8306	**0.7840***	0.8533	0.8922	0.8987

Table 10: Relative out-of-sample losses of HAR, HARQ, HAR_{log}, HARK, SHAR, SHARK, ARFIMA, HARST, $\text{HAR}^{\Delta J^{2\pm}}$ on S&P500 data for weekly ($j = 5$) and monthly ($j = 22$) forecasts in the period [Dec. 2002 - Jun. 2013]. We report losses computed both on a rolling window (RW) and on static window (SW). The presence of an asterisk indicates that the model is included in $\widehat{\mathcal{M}}_{90\%}$.

Electronic copy available at: https://ssrn.com/abstract=3089929

		φ_{het}	$\varphi^{(1)}_{\text{tvp}}$	$\varphi^{(2)}_{\text{tvp}}$	$\varphi^{(1)}_{\text{err}}$	$\varphi^{(2)}_{\text{err}}$
				j = 5		
				$\underline{M = 39}$		
MSE	RW	0.9302	0.9892	0.9640	0.9726	0.9477
	SW	0.9280	0.9833	0.9663	0.9615	0.9449
QLIKE	RW	0.8253	0.9731	0.9580	0.9755	0.9604
	SW	0.8176	0.9743	0.9563	0.9786	0.9605
				$\underline{M = 78}$		
MSE	RW	0.8853	0.9715	0.9757	0.9773	0.9815
	SW	0.8798	0.9684	0.9779	0.9762	0.9858
QLIKE	RW	0.7662	0.9631	0.9691	0.9726	0.9787
	SW	0.7625	0.9603	0.9673	0.9732	0.9803
				$\underline{M = 390}$		
MSE	RW	0.9190	0.9933	0.9942	0.9885	0.9894
	SW	0.9126	0.9918	0.9925	0.9885	0.9892
QLIKE	RW	0.8572	0.9728	0.9785	0.9830	0.9887
	SW	0.8530	0.9701	0.9701	0.9848	0.9848
				j = 22		
				$\underline{M = 39}$		
MSE	RW	0.8686	0.9842	0.9492	0.9595	0.9254
	SW	0.8675	0.9796	0.9498	0.9554	0.9263
QLIKE	RW	0.8023	1.0001	0.9072	0.9511	0.8627
	SW	0.7997	0.9949	0.9031	0.9489	0.8614
				$\underline{M = 78}$		
MSE	RW	0.8319	0.9677	0.9744	0.9662	0.9730
	SW	0.8258	0.9657	0.9786	0.9629	0.9758
QLIKE	RW	0.7697	0.9909	0.9358	0.9546	0.9014
	SW	0.7611	0.9913	0.9216	0.9707	0.9025
				$\underline{M = 390}$		
MSE	RW	0.9409	0.9990	0.9739	0.9822	0.9574
	SW	0.9325	0.9864	0.9535	0.9917	0.9587
QLIKE	RW	0.8612	0.9689	0.9425	0.9720	0.9455
	SW	0.8622	0.9633	0.9298	0.9780	0.9439

Table 11: Forecast gains for weekly ($j = 5$) and monthly ($j = 22$) forecast horizons. We denote by RW the forecast gains computed on a rolling window and by SW those computed on a static window.

Electronic copy available at: https://ssrn.com/abstract=3089929

	φ_{het}	φ_{err^1}	φ_{tvp^1}
	uSPA test		
t_{uSPA}	3.5273	2.8827	1.8295
p-value	0	0	0.0365

Table 12: MSE based uniform superior predictive ability test (uSPA) of Quaedvlieg (2019). We show the t_{uSPA} test statistics and related p-values based on a multi-horizon bootstrap with 5000 replications.

	$M = 39$	$M = 78$	$M = 390$
	β		
HARK	2.0595	2.9399	6.9714
	(0.2106)	(0.2856)	(2.7879)
SHARK	0.8969	1.2051	2.3594
	(0.2511)	(0.3036)	(1.8479)

Table 13: Maximum-likelihood estimates of the scaling parameter β in the dynamic specification 4. in Section (3.4). Standard errors are shown in parenthesis.

Electronic copy available at: https://ssrn.com/abstract=3089929

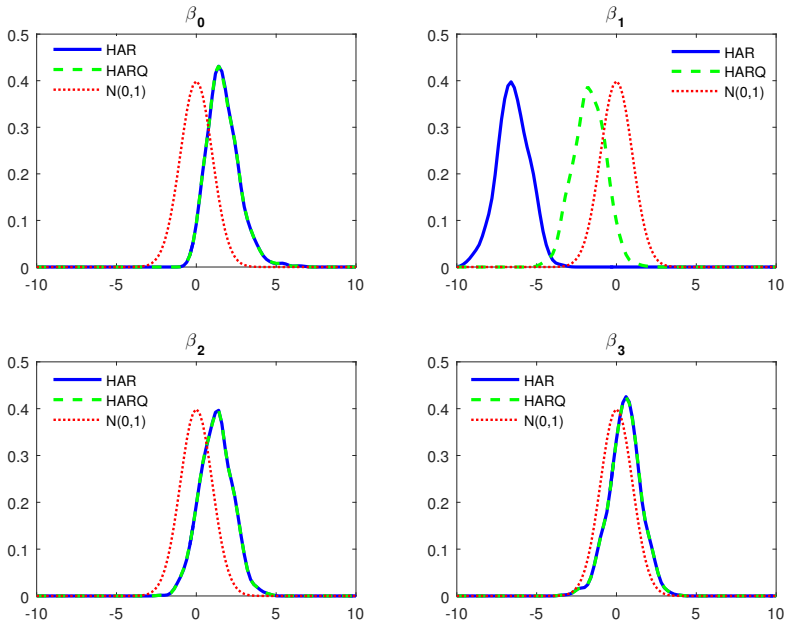

Figure 1: Kernel density estimates of standardized pivotal HAR and HARQ statistics estimated on $N = 1000$ Monte-Carlo simulations of $T = 1000$ observations of model (10), (11)

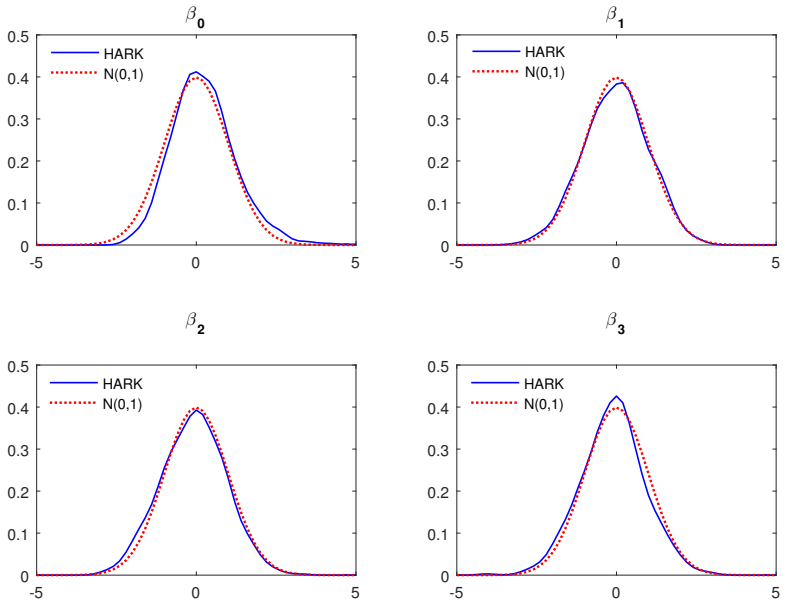

Figure 2: Kernel density estimates of standardized pivotal HARK statistics estimated on $N = 1000$ Monte-Carlo simulations of $T = 1000$ observations of model (10), (11)

Electronic copy available at: https://ssrn.com/abstract=3089929

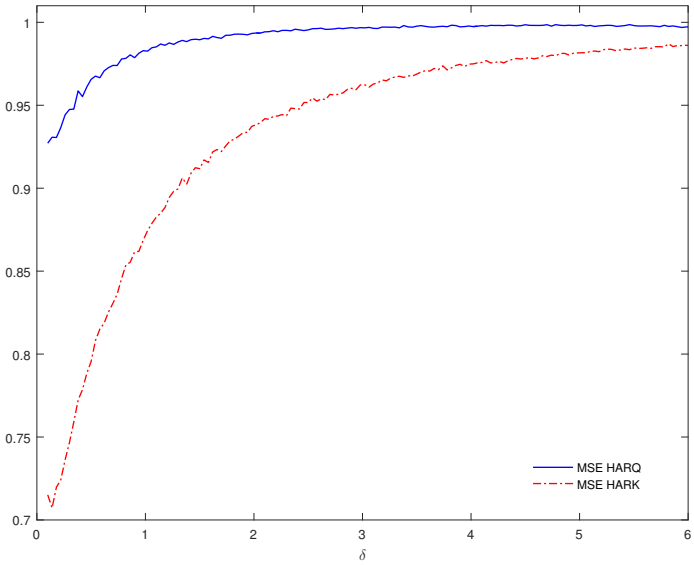

Figure 3: Out-of-sample MSE of HARQ and HARK models for different values of signal-to-noise ratio δ. Each MSE is divided by the MSE provided by the HAR model.

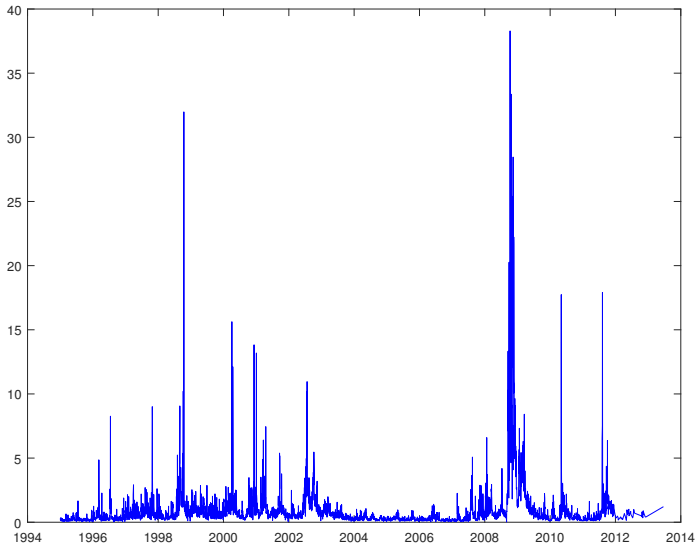

Figure 4: Time series of 5-min daily realized variance of S&P500 future prices from 03-01-1995 to 21-06-2013.

Electronic copy available at: https://ssrn.com/abstract=3089929

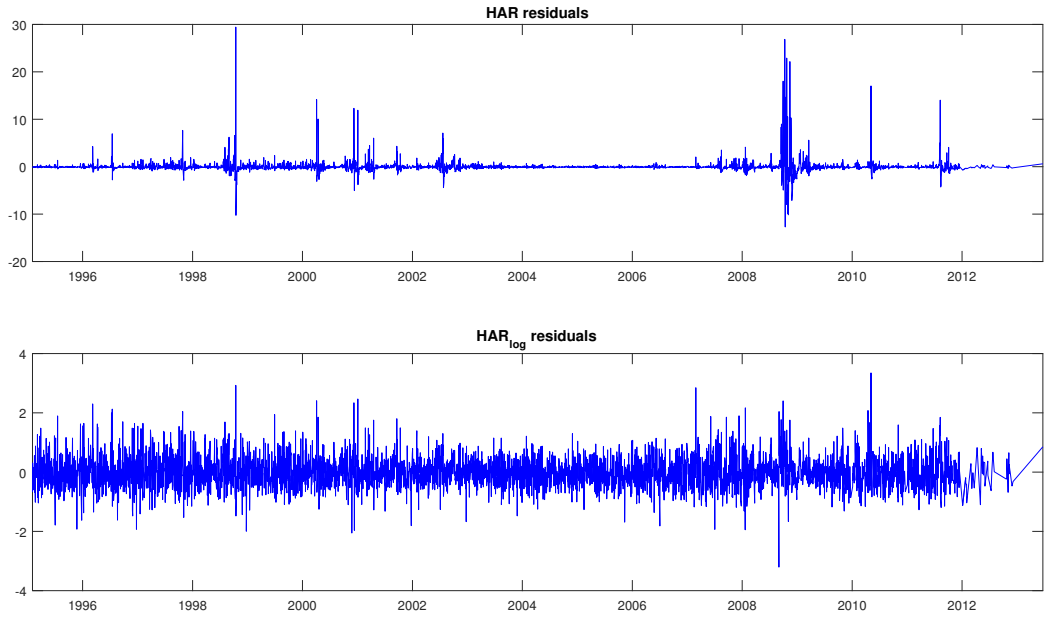

Figure 5: Residuals of HAR and $\text{HAR}_{\log}$ models estimated on the series of 5-min daily realized variance of S&P500 future prices from 03-01-1995 to 21-06-2013.

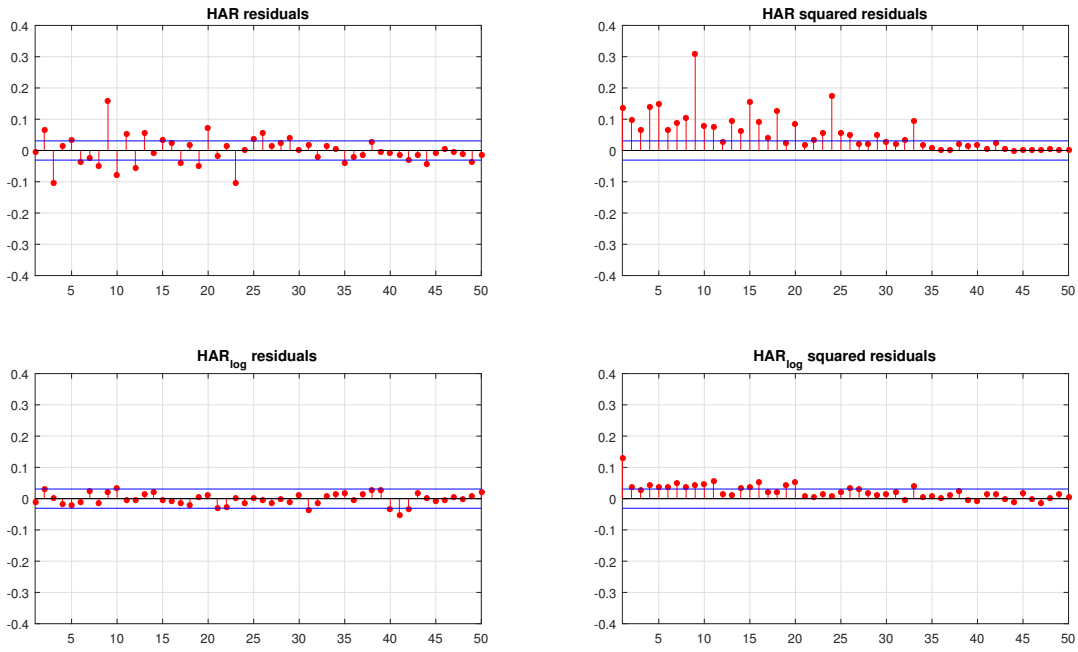

Figure 6: Sample autocorrelation function of HAR and $\text{HAR}_{\log}$ residuals and squared residuals estimated on the series of 5-min daily realized variance of S&P500 future prices from 03-01-1995 to 21-06-2013.

Electronic copy available at: https://ssrn.com/abstract=3089929

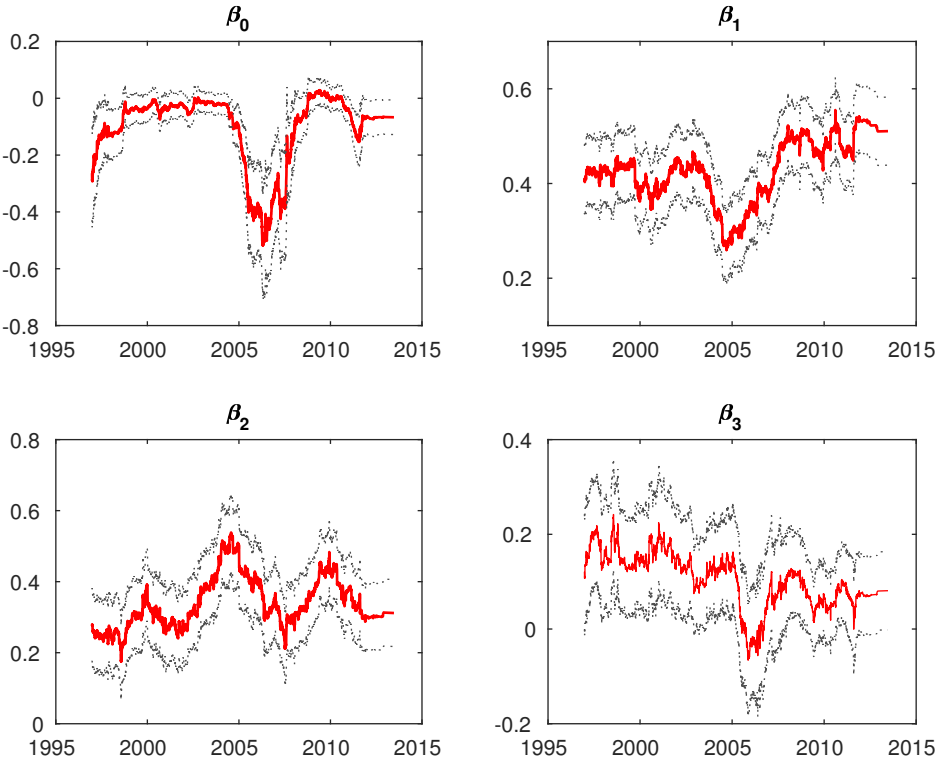

Figure 7: Dynamics of $\text{HAR}_{\log}$ coefficients obtained by estimating the model on a moving window of 1000 observations in the period 17-12-1998 − 21-06-2013. The red line denotes the estimated coefficient, while the two grey lines denote 95% confidence bands.

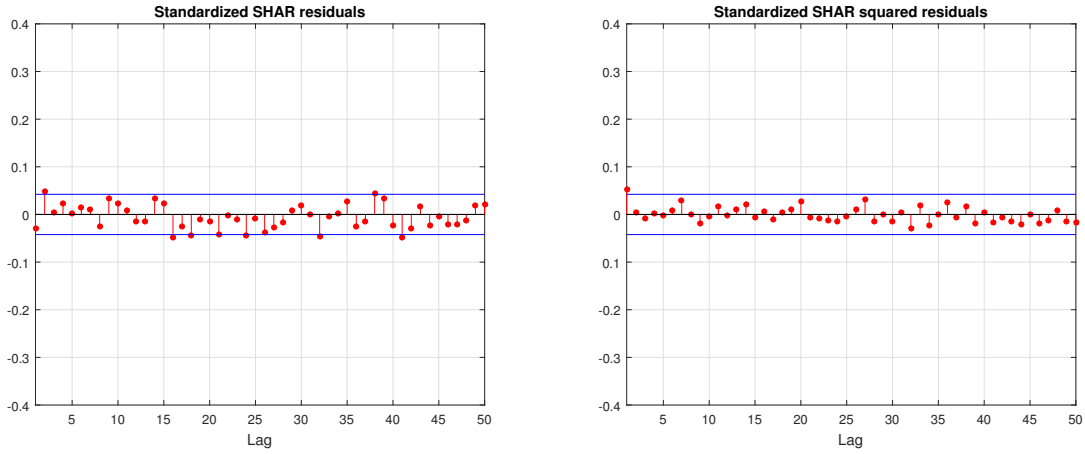

Figure 8: Sample autocorrelation function of SHAR standardized residuals and squared standardized residuals estimated the series of 5-min daily realized variance of S&P500 future prices from 03-01-1995 to 21-06-2013.

Electronic copy available at: https://ssrn.com/abstract=3089929

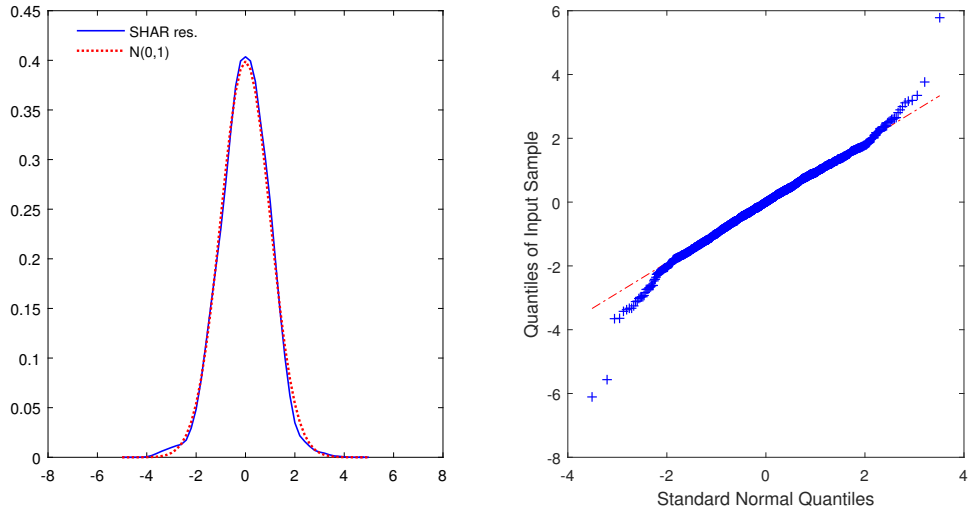

Figure 9: Kernel density estimate of SHAR standardized residuals computed on the series of 5-min daily realized variance of S&P500 future prices from 03-01-1995 to 21-06-2013. Q-Q plot of standardized residuals.

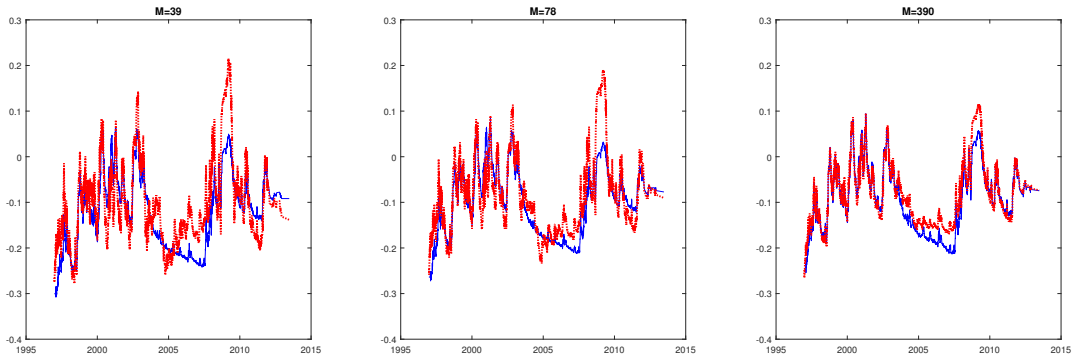

Figure 10: Dynamics of filtered SHAR (blue lines) and SHARK (red lines) $\beta_{0,t}$ parameter obtained by estimating the two models on S&P500 realized variance computed with $10, 5, 1$-min returns in the period 23-12-1996 − 21-06-2013.

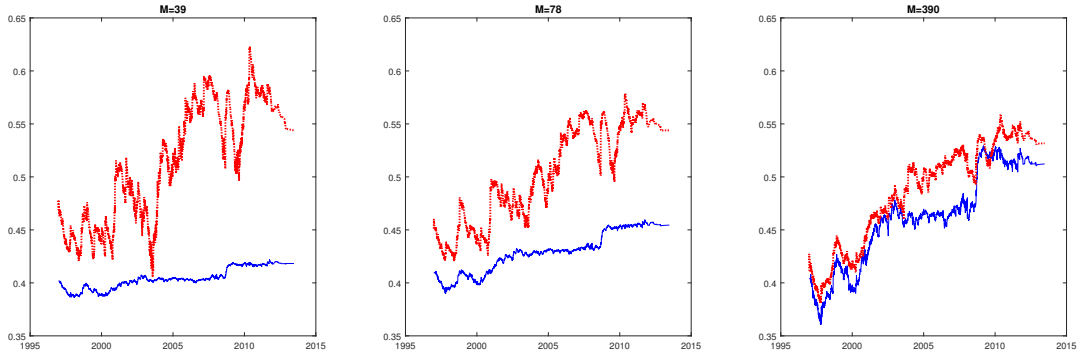

Figure 11: Dynamics of filtered SHAR (blue lines) and SHARK (red lines) $\beta_{1,t}$ parameters obtained by estimating the two models on S&P500 realized variance computed with $10, 5, 1$-min returns in the period 23-12-1996 − 21-06-2013.

Electronic copy available at: https://ssrn.com/abstract=3089929

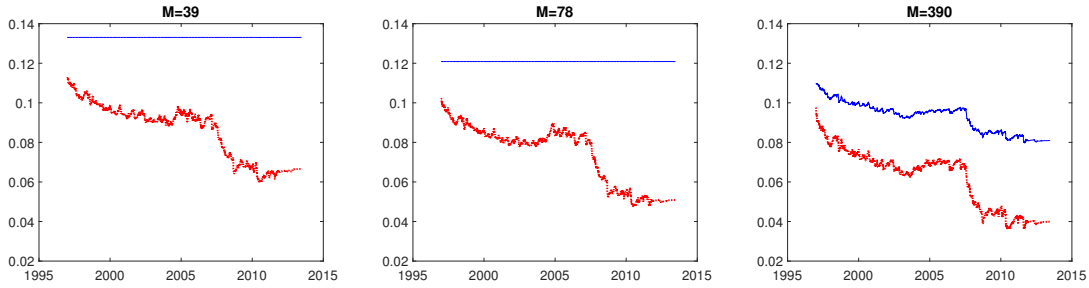

Figure 12: Dynamics of filtered SHAR (blue lines) and SHARK (red lines) $\beta_{3,t}$ parameters obtained by estimating the two models on S&P500 realized variance computed with $10, 5, 1$-min returns in the period 23-12-1996 − 21-06-2013.

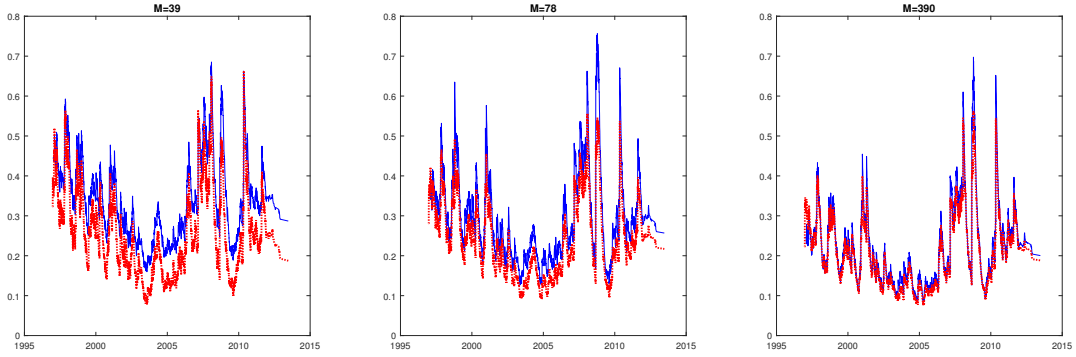

Figure 13: Dynamics of filtered SHAR (blue lines) and SHARK (red lines) q_t parameter obtained by estimating the two models on S&P500 realized variance computed with $10, 5, 1$-min returns in the period 23-12-1996 − 21-06-2013.

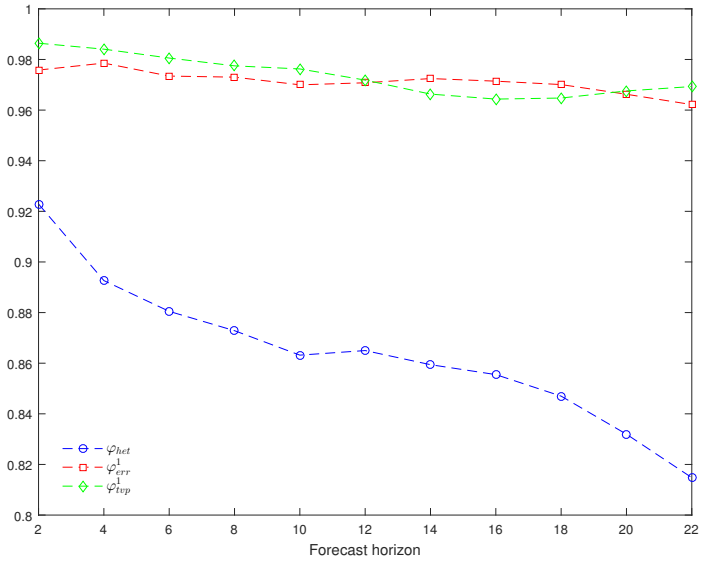

Figure 14: MSE-based forecast gains φ_{het}, φ^1_{err}, φ^1_{tvp} at different forecast horizons.

Electronic copy available at: https://ssrn.com/abstract=3089929