International Journal of Forecasting 39 (2023) 827–840

Contents lists available at ScienceDirect

International Journal of Forecasting

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

Time-varying variance and skewness in realized volatility measures☆

Anne Opschoor*, André Lucas

Vrije Universiteit Amsterdam and Tinbergen Institute, The Netherlands

ARTICLE INFO

Keywords:
Realized kernel
Heavy tails
Dynamic *F* distribution
Time-varying shape-parameters
Vol-of-vol
Score-driven dynamics

ABSTRACT

We propose new empirical models to capture the dynamics of the variance and skewness in realized volatility measures. We find that time-variation in variance and skewness of realized measures is a key empirical feature, even after accounting for well-known, stylized facts such as long-memory-type persistence and large incidental observations. Using a broad range of 89 US stocks across different sectors over 2001–2019, we show that these are not incidental phenomena of a few stocks but are widely shared. Accounting for dynamics in the variance and skewness of realized measures results in significantly better in-sample fit and out-of-sample unconditional density and quantile forecasts.

1. Introduction

Volatility is a key ingredient for asset pricing, volatility trading, and risk management; see for instance Sinclair (2013) and Caporin et al. (2017). Due to the availability of High-Frequency (HF) data, modern empirical volatility models use (direct) measurements of the daily volatility, such as the realized variance (Andersen & Bollerslev, 1998) or realized kernel (Barndorff-Nielsen et al., 2008). Examples include the HAR model of Corsi (2009), the MEM model of Engle and Gallo (2006), or the multivariate CAW model of Golosnoy et al. (2012). Also, hybrid models based on both returns and realized measures have been put forward, such as the univariate and multivariate HEAVY models of Shephard and Sheppard (2010) and Noureldin et al. (2012).

Stylized facts for realized volatility measures include long memory features, fat-tailedness, and a right skew. The fat-tailedness and right skew can be attributed to turbulent market periods and price jumps, even within the day (e.g. the Flash Crash in May 2011). Very few studies take the fat-tailedness of realized measures directly into account. Examples include Caporin et al. (2017) and Opschoor et al. (2018). Caporin et al. (2017) extend the MEM model of Engle and Gallo (2006) by including jumps under the assumption of a mixture of Gamma distributions for the realized variance to capture the skewed right tail of the volatility density. Opschoor et al. (2018) propose a matrix *F* distribution for realized covariance matrices, which is a continuous mixture of Wishart distributions. However, these models are static in terms of their variance (vol-of-vol) and skewness (skew-of-vol).

Intuitively, higher order moments of volatility such as the variance and skewness need not be constant over time. Relating volatility to the flow of information as in the seminal work of Andersen (1996), there is no ex-ante reason why the acceleration or deceleration of the flow of information should be constant over relatively calm versus more turbulent periods, such as the burst of the Internet bubble, the Global Financial Crisis, or the Sovereign debt crisis. Corradi et al. (2013) indeed show that volatility-of-volatility is time-varying and varies with the business cycle. Furthermore, Huang

☆ We appreciate the comments of the associate editor and referee, participants at the 11th Annual SoFiE Conference in Lugano, the NSEG meeting in Amsterdam, and seminar participants at the University of Parma and Vrije Universiteit Amsterdam that have all helped to improve the paper.

* Correspondence to: Vrije Universiteit Amsterdam, De Boelelaan 1105, 1081 HV, Amsterdam, The Netherlands.
E-mail address: a.opschoor@vu.nl (A. Opschoor).

https://doi.org/10.1016/j.ijforecast.2022.02.009
0169-2070/

et al. (2019) and Baltussen et al. (2018) document that volatility risk is time-varying and matters for asset pricing. Still, empirical models for realized volatility measures have been mainly static and failed to investigate the dynamics of the variance and skewness and the possible time-varying acceleration of the information flow.

This paper develops new empirical models for volatility that allow the shape parameters and thus the implied variance and skewness of volatility to be time-varying. By modeling these two moments implicitly via the distribution's parameters, the model remains applicable even when the (implied) higher order moments such as skewness do not exist. The goal is to create more flexibility in the variance (also abbreviated as the vol-of-vol) and the volatility distribution's skewness patterns, thus allowing the information flow to accelerate or decelerate over time. We measure volatility through the realized kernel of Barndorff-Nielsen et al. (2008). We model the dynamic behavior of the kernel using a score-driven time-series model, allowing simultaneously for time-variation in the mean volatility level and the shape parameters; see Creal et al. (2013). To the best of our knowledge, time-varying tail shape parameters have been rarely studied before, and then only for *return* distributions rather than for (realized) volatilities (see for instance Gerlach et al., 2013; Lucas & Zhang, 2016). Our approach has a closed-form expression for the likelihood function, allowing straightforward estimation and inference by maximum likelihood. In addition, as the method is observation-driven in the classification of Cox (1981), we can easily allow standard stylized facts into the model, such as the long-memory-type persistence of volatility using the HAR approach of Corsi (2009). Combined with the optimality results for score-driven models of Blasques et al. (2015), our framework can thus substantially reduce the Kullback–Leibler divergence between the model and the unknown data generating process for realized volatility dynamics.

Our model for realized kernels builds on a scaled F distribution, where we impose score-driven dynamics for the time-varying mean as well as for the two degrees of freedom (DoF) parameters. The fat tail of the F distribution appears to substantially improve the fit of the model to empirical time-series of realized kernels; see, for instance, the matrix version of this distribution in Opschoor et al. (2018). Although the variance and skewness of the F distribution depend on *both* DoF parameters, we show that the second DoF parameter mostly affects the tail shape (skewness) of the distribution, while the first DoF mostly affects the dispersion (or variance). Evidence that this dispersion or vol-of-vol can be time-varying is found in, for instance, Corsi et al. (2008).

Our empirical application uses the new model to describe daily realized kernels of a broad range of 89 US stocks across industries from January 2001 to December 2019. In the sample, the statistical fit increases significantly when allowing for time-varying DoF parameters compared to benchmarks such as the MEM model (with HAR dynamics) of Engle and Gallo (2006), the univariate GAS F model with HAR dynamics of Opschoor et al. (2018), or model specifications for the logarithm of the realized kernel. The latter are interesting strong benchmark models, as modeling the *logarithm* of the realized kernel immediately reduces the fat-tailedness of the distribution; see, for instance, the HAR log volatility of Corsi (2009), the Realized log GARCH model of Hansen et al. (2012), or the multivariate VARFIMA model used by Chiriac and Voev (2011).

Out-of-sample, we assess the economic significance of our results by considering 1-step ahead density forecasts and Volatility-at-Risk (VolaR) predictions. Our new model strongly outperforms all other models in terms of unconditional density forecasts: it is part of the model confidence set (Hansen et al., 2011) in 87 of the 89 cases. As a comparison, the MEM model with the Gamma distribution enters the set only 11 times. The results strongly support that it is important to allow time-varying shape parameters to produce an adequate unconditional VolaR. The time-variation in the variance and skewness of volatility is widely shared and does not appear to be limited to just a few assets. In terms of conditional density forecasts and conditional VolaR predictions, all models considered in this paper face challenges. The score-driven models proposed in this paper require more than one signal to disentangle fat-tailedness from increases in volatility, which results in a first order autocorrelation and rejection of standard tests for conditional coverage. However, the autocorrelation is short-lived and disappears already at the second partial autocorrelation. By contrast, the standard conditional tests mask some of the more problematic behavior of thin-tailed models: the resulting transformed probability integral transforms (PITs) exhibit sequences of influential observations that induce under-rejection of serial correlation tests. We discuss these features in detail.

The rest of this paper is set up as follows. In Section 2, we introduce the new model for the dynamic shape parameters of the realized kernel distribution, as well as its competing benchmarks. In Section 3, we give a brief overview of the data used and cover the empirical application. We conclude in Section 4.

2. Modeling realized volatility

2.1. The model for static shape parameters

As our measure of realized volatility $RK_t \in \mathbb{R}$ for days $t = 1, \ldots, T$, we use the realized kernel as computed following Barndorff-Nielsen et al. (2009). To allow for fat-tailed behavior of RK_t, we follow Opschoor et al. (2018) and assume RK_t follows a conditional F distribution:

$$p(RK_t|\mu_t, \mathcal{F}_{t-1}; \nu_1, \nu_2) = \frac{\Gamma((\nu_1+\nu_2)/2)}{\Gamma(\nu_1/2)\Gamma(\nu_2/2)} \frac{RK_t^{(\nu_1-2)/2} \left(\frac{\nu_1}{\mu_t(\nu_2-2)}\right)^{\nu_1/2}}{\left(1+\frac{\nu_1}{\nu_2-2}\frac{RK_t}{\mu_t}\right)^{(\nu_1+\nu_2)/2}}, \quad (1)$$

where $\mathcal{F}_{t-1}$ denotes the information set containing all realized kernels up to and including time $t-1$, μ_t contains the time-varying mean of RK_t, and ν_1 and ν_2 are the degrees of freedom (DoF) parameters. We assume that $\nu_2 > 2$, such that the conditional mean $\mathbb{E}[RK_t|\mathcal{F}_{t-1}] = \mu_t$

exists.[1] When $\nu_2 \to \infty$, the F distribution collapses to the χ^2 or Wishart distribution. By an appropriate choice of μ_t, this also covers the Gamma distribution used by Engle and Gallo (2006) and Caporin et al. (2017), among others.

The corresponding (conditional) variance and skewness of the F distribution are given by

$$\text{Var}[RK_t|\mathcal{F}_{t-1}] = 2\,\frac{(\nu_1+\nu_2-2)}{\nu_1(\nu_2-4)}\,\mu_t^2, \tag{2}$$

$$\text{Skew}[RK_t|\mathcal{F}_{t-1}] = \frac{(\nu_2-2)^3(2\nu_1+\nu_2-2)\sqrt{8(\nu_2-4)}}{\nu_2^3(\nu_2-6)\sqrt{\nu_1(\nu_1+\nu_2-2)}}\,\mu_t^3, \tag{3}$$

where the conditional variance and skewness exist if $\nu_1 > 0$, and if $\nu_2 > 4$ (vol-of-vol) and $\nu_2 > 6$ (skewness), respectively. Note that even if the conditional skewness (or vol-of-vol) does not exist, the model remains applicable if $\nu_2 > 2$ given we estimate ν_1 and ν_2 and only treat vol-of-vol and skewness as implied by the underlying model parameters. This is important, as the existence of higher order moments for high-frequency data may sometimes be problematic.

Eqs. (2) and (3) show that both the variance and skewness depend on μ_t, i.e., on the conditional expected level of RK_t. Hence by allowing for a time-varying level μ_t of the volatility distribution, the skewness and kurtosis change mechanically as well, even if the DoF parameters are constant through time. Furthermore, both the conditional variance and skewness of the F distribution depend on ν_1 *and* ν_2. Therefore, it is not immediately apparent how these parameters affect each of these quantities separately. To disentangle these effects, Fig. 1 shows a surface plot for the variance and skewness for different combinations of ν_1 and ν_2 while keeping μ_t constant. We obtain two important insights. First, the variance is decreasing in ν_2 and ν_1. However, if both values become small, the variance increases relatively more due to ν_1 compared to ν_2. Second, the skewness is decreasing in both ν_2 and ν_1, but for small values of ν_1 and ν_2, the impact of ν_2 on the skewness is larger than the impact of ν_1. We, therefore, label ν_2 as the 'tail shape' or 'skewness' parameter and ν_1 as the 'dispersion' or 'vol-of-vol' parameter.

Our modeling framework starts with the univariate version of the multivariate HEAVY GAS model of Opschoor et al. (2018) endowed with the HAR dynamics for μ_t to accommodate the long-memory-type persistence of the realized kernels; see Corsi (2009). The dynamics of the volatility process μ_t follow the generalized autoregressive score (GAS) framework of Creal et al. (2013). We obtain

$$RK_t = \mu_t \epsilon_t, \qquad \epsilon_t \sim F(1, \nu_1, \nu_2), \tag{4}$$

$$\mu_{t+1} = \omega_\mu + \alpha_\mu\, s_{\mu,t} + \beta_{\mu,1}\,\mu_{l_1,t} + \beta_{\mu,2}\,\mu_{l_2,t} + \beta_{\mu,3}\,\mu_{l_3,t}, \tag{5}$$

$$s_{\mu,t} = S_t \nabla_t = S_t \frac{\partial \log p(RK_t|\mu_t, \mathcal{F}_{t-1}; \nu_1, \nu_2)}{\partial \mu_t} = \frac{\nu_1}{\nu_1+1}\left(\frac{\frac{\nu_1+\nu_2}{\nu_2-2}RK_t}{(1+\frac{\nu_1}{\nu_2-2}\frac{RK_t}{\mu_t})} - \mu_t\right), \tag{6}$$

where $\partial \log p_{RK}(RK_t|\mu_t, \mathcal{F}_{t-1}; \nu_1, \nu_2)/\partial \mu_t$ denotes the score with respect to μ_t, S_t is a scaling factor, $\mu_{l,t} = l^{-1}\sum_{i=0}^{l-1}\mu_{t-i}$, and $l_1 = 1, l_2 = 12$, and $l_3 = 60$.[2] Following Opschoor et al. (2018), we scale the score by $2\mu_t^2/(\nu_1+1)$ to account for the curvature of $\log p_{RK}(RK_t|\mu_t, \mathcal{F}_{t-1}; \nu_1, \nu_2)$ with respect to μ_t. This quantity is proportional to the inverse conditional Fisher information with respect to μ_t. In the remainder of this paper, we label this model the GAS-HAR model.

The scaled score (6) has an appealing interpretation for the dynamics of μ_t. It is the difference between a weighted value of RK_t and μ_t, where the weight accounts for the fat-tailedness of RK_t. That is, large values of RK_t imply a low weight, such that the impact of a large RK_t on μ_{t+1} will be mitigated. Fig. 2 visualizes this property of the score. The figure can be interpreted as a news impact curve. Small values of RK_t decrease the expected value μ_t. As RK_t increases, the score increases as well, but the increase levels off and becomes flat for large values of RK_t, refraining μ_{t+1} from increasing unboundedly as long as ν_2 is finite.

Also, note that the existence of higher order moments for high-frequency data may be problematic. The current framework with the F distribution and GAS dynamics accommodates this more easily than standard MEM type dynamics. In particular, note that the score-driven recursion (6) does not consider RK_t directly, but rather a weighted version. This weighted version is uniformly bounded in RK_t. This considerably relaxes the conditions

[1] We have chosen to parameterize the distribution such that μ_t coincides with the mean; hence the division by $\nu_2 - 2$ and the requirement that the mean exists. We could make the model even more flexible by not requiring the first moment to exist. That would replace the divisions by $\nu_2 - 2$ (twice) in the density expression by ν_2 and consequently slightly change the density expression and subsequent scores. We would then only require $\nu_2 > 0$ rather than $\nu_2 > 2$. The parameter μ_t would become a generic scale parameter rather than the expectation of RK_t. As the existence of variances in financial applications is mostly accepted, we stick to the current parameterization for more straightforward interpretability.

[2] Here we deviate from the original specification of Corsi (2009) for two reasons. First, we put the HAR dynamics on μ_t rather than on the score $s_{\mu,t}$. This makes sense as the scores are a martingale difference sequence. The difference is also smaller than one might think at first sight once we note that the original Corsi model can be written as

$$\begin{aligned} RK_t &= \omega + \beta_1 RK_{l_1,t} + \beta_2 RK_{l_2,t} + \beta_3 RK_{l_3,t} + e_t, \\ &= \omega + \beta_1\mu_{l_1,t} + \beta_2\mu_{l_2,t} + \beta_3\mu_{l_3,t} + \alpha_1(RK_{l_1,t} - \mu_{l_1,t}) \\ &\quad + \alpha_2(RK_{l_1,t} - \mu_{l_2,t}) + \alpha_3(RK_{l_1,t} - \mu_{l_3,t}) + e_t. \end{aligned}$$

where $\alpha_j = \beta_j$. The right-hand side now contains the lag polynomial on μ_t as in (5) and three martingale differences $RK_{l_j,t} - \mu_{l_j,t}$, the first of which also enters the right-hand side of (5). The remaining two have variances which are a factor l_1/l_2 and l_1/l_3 lower due to the averaging operators, and are therefore ignored in (5). An empirical check confirms that the fit hardly improves if these last two martingale differences are added to the model specification. Finally note that RK_t and e_t on the left-hand and right-hand side of the Corsi model have a similar effect as the absence of an additional error term in (5) combined with the conditional F assumption in (4). Our choice of l_1, l_2, l_3 follows Jin and Maheu (2016), who also apply the HAR dynamics on the filtered (co)variance and empirically find $l_2 \approx 13$ and $l_3 = 60$. We confirmed empirically that our full-sample fit increases when using $l_1 = 1, l_2 = 12$, and $l_3 = 60$ instead of the original values suggested by Corsi (2009). The methodology could, of course, be further extended by introducing a data-driven way to select the averages to include in the transition Eq. (5) as in for instance (Audrino & Knaus, 2016).

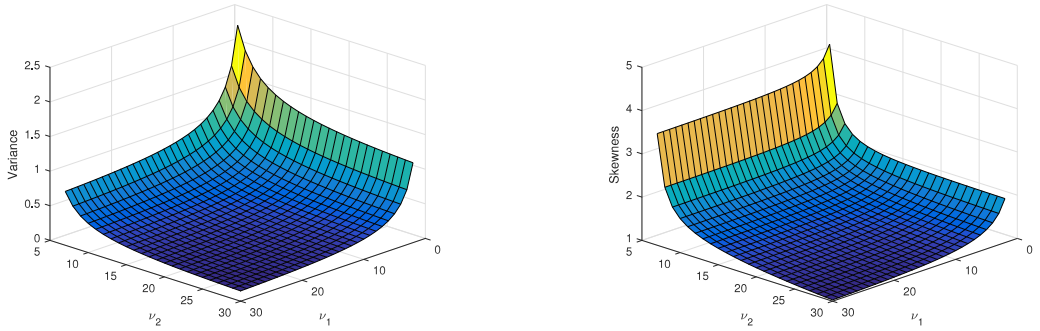

Fig. 1. Variance and Skewness of the F distribution. This figure shows the variance and skewness of an F distributed random variable with unit mean as a function of ν_1 and ν_2.

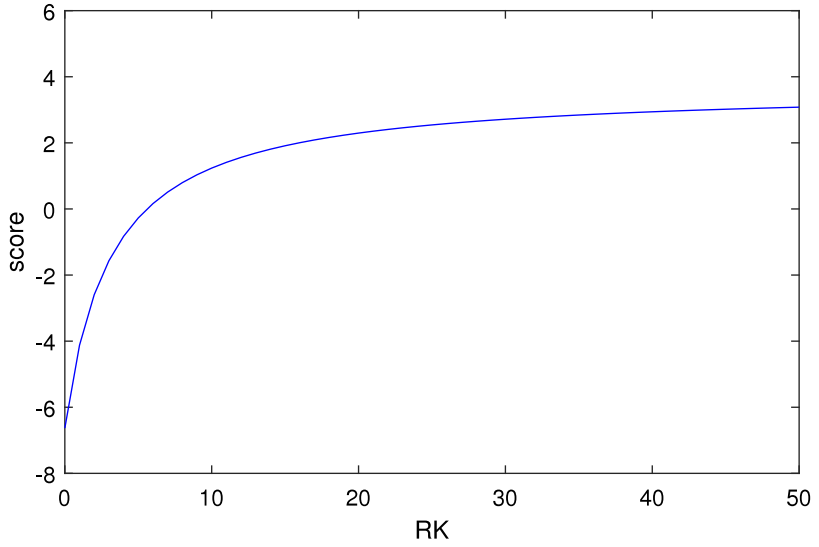

Fig. 2. Score of the F distribution with respect to μ_t. This figure shows the score of the F distribution (given in Eq. (6)) with respect to μ_t for various values of RK. The score is computed by setting $\nu_1 = 18$, $\nu_2 = 10$ and μ_t to 7.

needed to ensure proper convergence of the recursion and the existence of moments of μ_t, even if RK_t itself is very fat-tailed with $1 < \nu_2 < 2$; see also (Blasques et al., 2022). This stands in sharp contrast with usual MEM dynamics, which depend on unweighted versions of RK_t and typically require much stricter conditions for good asymptotic properties.

2.2. A dynamic model for shape parameters

The flexibility of the GAS framework is that we can easily handle time-variation in other parameters than μ_t. This includes parameters describing the skewness and variance of ϵ_t. As the introduction argues, we can also expect time-variation in parameters like ν_1 and/or ν_2. For instance, one might expect information gathering to accelerate much more quickly during crisis periods than during calm periods. In terms of the parameters, this means that we might expect the conditional tail shape of RK_t to become relatively fat during turbulent times, while in calm periods RK_t might have lighter conditional tails, implying time-variation in ν_2.

We consider the following two time-varying parameter models for the DoF parameters. In the first model, we allow for a time-varying dispersion parameter $\nu_{1,t}$,

$$
\begin{aligned}
f_{1,t+1} &= \omega_{\nu_1} + \alpha_{\nu_1} s_{\nu_1,t} + \beta_{\nu_1} f_{1,t}, \\
\nu_{1,t} &= 2 + \exp(f_{1,t}), \qquad (7a) \\
s_{\nu_1,t} &= \frac{\partial \log p(RK_t|\mu_t, \mathcal{F}_{t-1}; \nu_{1,t}, \nu_2)}{\partial \nu_{1,t}} \frac{\partial \nu_{1,t}}{\partial f_{1,t}} \\
&= \left[\frac{1}{2}\psi\left(\frac{\nu_{1,t}+\nu_2}{2}\right) - \frac{1}{2}\psi\left(\frac{\nu_{1,t}}{2}\right) \right. \\
&\quad + \frac{1}{2}\left(\log\left(\frac{\nu_{1,t}}{\nu_2 - 1}\right) + 1\right) + \frac{1}{2}\log\left(\frac{RK_t}{\mu_t}\right) - \\
&\quad \frac{1}{2}\log\left(1 + \frac{\nu_{1,t}}{\nu_2 - 2}\frac{RK_t}{\mu_t}\right) \\
&\quad \left. - \frac{\nu_{1,t}+\nu_2}{2} \frac{\frac{1}{\nu_2 - 2}\frac{RK_t}{\mu_t}}{(1 + \frac{\nu_{1,t}}{\nu_2 - 2}\frac{RK_t}{\mu_t})} \right] (\nu_{1,t} - 2), \qquad (7b)
\end{aligned}
$$

where ψ denotes the digamma function $\psi(x) = \partial \log \Gamma(x)/\partial x$. We label this the GAS-HAR-ν_1 model. As we have seen before, this model focuses on modeling

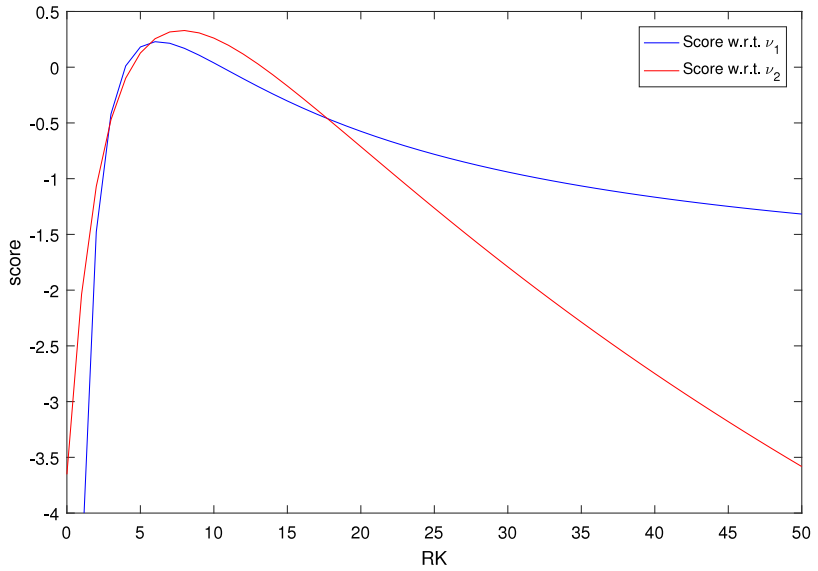

Fig. 3. Score of the F distribution with respect to ν_1 and ν_2. This figure shows the score of the F distribution (given in Eq. (8b)) with respect to ν_1 (blue line) and ν_2 (red line) for various values of RK. The former (latter) score is computed by setting $\nu_1 = 17\,(19)$, $\nu_{2,t} = 13\,(17)$ and μ_t to 7. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

the vol-of-vol (see also Corsi et al., 2008). The second model considers time-variation in the tail shape via the parameter ν_2. We have

$$f_{2,t+1} = \omega_{\nu_2} + \alpha_{\nu_2} s_{\nu_2,t} + \beta_{\nu_2} f_{2,t}, \quad \nu_{2,t} = 2 + \exp(f_{2,t}), \tag{8a}$$

$$\begin{aligned} s_{\nu_2,t} &= \frac{\partial \log p(RK_t|\mu_t, \mathcal{F}_{t-1}; \nu_1, \nu_{2,t})}{\partial \nu_{2,t}} \frac{\partial \nu_{2,t}}{\partial f_{2,t}} \\ &= \left[\frac{1}{2}\psi\left(\frac{\nu_1+\nu_{2,t}}{2}\right) - \frac{1}{2}\psi\left(\frac{\nu_{2,t}}{2}\right) - \frac{1}{2}\frac{\nu_1}{\nu_{2,t}-1} \right. \\ &\quad - \frac{1}{2}\log\left(1 + \frac{\nu_1}{\nu_{2,t}-2}\frac{RK_t}{\mu_t}\right) + \\ &\quad \left. \frac{\nu_1+\nu_{2,t}}{2} \frac{\frac{\nu_1}{(\nu_{2,t}-2)^2}\frac{RK_t}{\mu_t}}{(1+\frac{\nu_1}{\nu_{2,t}-2}\frac{RK_t}{\mu_t})} \right] (\nu_{2,t}-2). \end{aligned} \tag{8b}$$

The parameterization $\nu_{2,t} = 2 + \exp(f_{2,t})$ ensures that $\nu_{2,t} > 2$ for all $f_{2,t} \in \mathbb{R}$, such that the mean of RK_t always exists.[3] We label this model the GAS-HAR-ν_2. Note that we can easily combine models (7a)–(7b) and (8a)–(8b) into a model with both time-varying mean, vol-of-vol, and tail shape. We call this combined model the GAS-HAR-ν_{1-2} model.

Even though $\nu_{1,t}$ and $\nu_{2,t}$ relate to the higher order moments of the conditional distribution, the score expressions in Eqs. (7b) and (8b) show that the dynamics of $f_{1,t}$ and $f_{2,t}$ are *not* driven by high order powers of RK_t such as RK_t^k for $k \geq 3$. A similar result is found by Lucas and Zhang (2016), who model the tails of asset returns by varying the degrees of freedom parameter of a (skewed) t distribution. The fact that the time-varying parameters $\nu_{1,t}$ and $\nu_{2,t}$ react to RK_t logarithmically rather than as a power function lends this model considerable stability, particularly for noisy data; see also Blasques et al. (2022) for the stability of score models. Also, note that we model $\nu_{1,t}$ and $\nu_{2,t}$ rather than the vol-of-vol and the skew-of-vol. This implies that when $\nu_{2,t}$ falls below six or even four, the skew-of-vol respectively vol-of-vol may no longer exist. The failure of these moments to exist at certain times, however, does not invalidate the validity or stability of the recursions in (7b) or (8b), as these only require a logarithmic moment of RK_t. Also, the time-variation in $\nu_{1,t}$ and $\nu_{2,t}$ remains valid, even though the higher order moments *implied* by these parameters may not exist at particular time points.

Fig. 3 visualizes the impact of the score on both $\nu_{1,t}$ and $\nu_{2,t}$ by varying RK_t. The figure clearly shows that when RK becomes large, both scores become more negative, and hence the values of $\nu_{1,t}$ and $\nu_{2,t}$ decrease. Note that the impact of large values of the realized kernel on $\nu_{2,t}$ is much more significant than for $\nu_{1,t}$. This is in line with Fig. 1, which confirms our finding that $\nu_{2,t}$ is more related to the tail of the distribution and changes more than the vol-of-vol related parameter $\nu_{1,t}$ for large RK_t. However, the *rate* of change is much more moderate than compared to a power RK_t^k of RK_t for $k \geq 3$. As mentioned, this robustness feature is essential for the time-series dynamics of $\nu_{1,t}$ and $\nu_{2,t}$, which will react much less violently to large incidental observations.

Fig. 1 also shows that small values (near zero) of RK_t result in lower values of $\nu_{1,t}$ and $\nu_{2,t}$. The reason is that for low values of these parameters, the F distribution also exhibits more leptokurtosis, i.e., a higher peak near zero, coinciding with more realizations of RK_t close to zero.

The static parameters in all models can be estimated by maximum likelihood using a standard prediction error decomposition. We first estimate the benchmark GAS model with only μ_t time-varying and ν_1 and ν_2 constant.

[3] Alternatively, one could set $\nu_{2,t} = 6 + \exp(f_{2,t})$ to ensure skewness always exists. As argued earlier, such a restriction is unnecessary for the GAS-HAR model.

Next, we make $\nu_{1,t}$ and/or $\nu_{2,t}$ time-varying. We parameterize the intercept ω for the GAS-HAR models as $(1-\beta)\cdot\bar{f}$, where $\bar{f}$ denotes unconditional mean of the time-varying parameter f_t, and estimate $\bar{f}$ rather than ω.

2.3. Modeling the logarithm of the realized kernel

A different route to model the realized kernel volatility is to propose model specifications for the logarithm of the realized kernel. Examples of log-specifications for volatility are Corsi (2009), Hansen et al. (2012) and Chiriac and Voev (2011), amongst others. Modeling the logarithm has some advantages, such as off-setting large observations by the log operation, ensuring positivity of the variance, and cutting the right-skewness of the realized kernel distribution. In addition, a model for the log volatility using a more standard normal density for the innovations could arguably be labeled as a simpler alternative to a model for RK_t directly using the more non-standard F distribution.

To endow the model for log volatility with sufficient dynamic flexibility, we use a score-driven time-varying mean (again including HAR dynamics to account for the long-memory-type persistence of the log volatility process) and vol-of-vol, following Corsi et al. (2008). The model is given by

$$\begin{aligned}
\log RK_t &= \mu_t + \epsilon_t = \mu_t + \sigma_t u_t, \qquad u_t \sim \mathrm{N}(0,1), && \text{(9a)}\\
\mu_{t+1} &= \omega_\mu + \alpha_\mu s_{\mu,t} + \beta_{\mu,1}\,\mu_{l_1,t} + \beta_{\mu,2}\,\mu_{l_2,t} \\
&\quad + \beta_{\mu,3}\,\mu_{l_3,t}, \\
\sigma^2_{t+1} &= \omega_\sigma + \alpha_\sigma s_{\sigma^2,t} + \beta_\sigma \sigma^2_t, && \text{(9b)}\\
s_{\mu,t} &= S_{\mu,t}\frac{\partial \log p(\log RK_t|\mu_t, \mathcal{F}_{t-1}; \nu_0)}{\partial \mu_t} = \epsilon_t - \mu_t, && \text{(9c)}\\
s_{\sigma^2,t} &= S_{\sigma^2,t}\frac{\partial \log p(\log RK_t|\mu_t, \sigma^2, \mathcal{F}_{t-1}; \nu_0)}{\partial \sigma^2_t} \\
&= \epsilon^2_t - \sigma^2_t, && \text{(9d)}
\end{aligned}$$

where σ^2_t denotes the vol-of-vol, and $S_{\mu,t}$ and $S_{\sigma^2,t}$ are the conditional inverse Fisher information matrices. We also estimate a restricted version of this model with static vol-of-vol $\sigma^2_t = \sigma^2$. The models are labeled GAS-HAR-log-N-VoV and GAS-HAR-log-N for the model with dynamic and static vol-of-vol, respectively. Note that we assume a conditional Normal distribution for u_t. This implies that $\log RK_t$ is also conditionally Normal distributed. Subsequently, RK_t follows a log-Normal distribution with known analytical expressions for the probability density function (pdf) and cumulative density function (cdf). Of course, the model can also allow for fat-tails by assuming, for instance, a conditional Student's t distribution for $\log RK_t$. However, this would imply that no integer conditional moments exist for RK_t itself, as it would be the exponential of a student's t random variable. Since the existence of the first moments of RK_t is key in our forecasting procedure, we do not incorporate such generalizations here and stick to the normal distribution. Note that this problem does not occur in our new GAS-HAR-F: as we model RK_t directly as having a conditional F distribution, first, second, and third conditional moments exist if $\nu_2 > 2, 4, 6$, respectively, which follows immediately from Eqs. (2) and (3).

3. Empirical application

In this section, we first provide a brief overview of the data. We then present an in-sample analysis by estimating our models using the whole sample to assess the possibly time-varying tail behavior of the realized kernels. Finally, we show the forecasting power of the new models by computing 1-step-ahead density and Volatility-at-Risk (VolaR) forecasts.

3.1. Data

The data consist of daily realized kernels of 89 U.S. equities from January 2, 2001, until December 31, 2019. All stocks are chosen across a broad set of different sectors and belong to the top 10%–15% w.r.t. Dec 2019 market capitalization within their sector.

Table 1 provides an overview of the individual stocks. We have $T = 4,713$ trading days after leaving out days where the market was open for half a day. We retrieve consolidated trades (transaction prices) from the Trade and Quote (TAQ) database from 9:30 until 16:00. Before 2015, the time-stamp precision equals one second; after 2015, it increases to one millisecond. After cleaning the high-frequency data following the guidelines of Barndorff-Nielsen et al. (2009) and Brownlees and Gallo (2006), we construct realized kernels based on 5-minute returns following again Barndorff-Nielsen et al. (2009).

Fig. 4 shows the evolution of RK_t for four random companies: Fluor Corp, KeyCorp, Boston Properties INC, and Consol Energy Inc. There are quite some peaks in the data for all stocks, especially during the 2007 Global Financial Crisis (GFC). The implication is that the distribution of the realized kernel might be considerably right-skewed. Fig. 5 illustrates this by plotting the sample skewness of all 89 stocks using the full sample. For many stocks, the skewness exceeds ten or more. Such values are typically caused by incidental very large values of RK_t, which can disrupt the estimated dynamic pattern of the realized kernel (in particular, μ_t) if not properly accounted for.

3.2. In-sample results

Using the full sample 2001–2019, we estimate four different GAS-HAR models for the realized kernel: a benchmark model with two static DoF parameters and models where either ν_1 or ν_2 or both vary over time. In addition, we augment the MEM model of Engle and Gallo (2006) with HAR dynamics to provide another competitive benchmark. The MEM model with HAR dynamics on RK_t is given by

$$\begin{aligned}
RK_t &= \mu_t\epsilon_t, \qquad \epsilon_t \sim \mathrm{Gamma}(1, \nu_1), \\
\mu_{t+1} &= \omega_\mu + \alpha_\mu RK_t + \beta_{\mu,1}\mu_t + \beta_{\mu,2}\,RK_{l_2,t} + \beta_{\mu,3}\,RK_{l_3,t},
\end{aligned} \qquad \text{(10)}$$

with $RK_{l,t} = l^{-1}\sum_{i=0}^{l-1} RK_{t-i}$, and $l_2 = 12$, and $l_3 = 60$, thus using the same lag lengths as for the GAS-HAR models. Finally, we estimate two models for the logarithm of the realized kernel as discussed in Section 2.3: the GAS-HAR-log-N with static and time-varying variance parameters, respectively.

Table 1
S&P 500 constituents.

Sec Nr.	Sector	# Comp.	Tickers
11	Agriculture	1	WY
21	Mining, oil and gas extraction	6	HAL, PXD, SU, CNX, SLB, OXY
22	Utilities	5	AES, AEP, AEE, DUK, SO
31–33	Manufacturing	34	KO, MO, GIS, CPB, LLY, PFE XOM, ABT, JNJ, MUR, CVX, MRK, BMY IP, CL, AVP, EL, LPX, MDT, AA BA, CAT, GE, BAX, DOV, F, NOC HPQ, TSM, UTX, A, IR, GD, ATI
42	Wholesale trade	4	HON, PG, SYY, MMM
44–45	Retail trade	6	HD, ANF, WSM, JCP, WMT, TGT
48	Transportation	4	WMB, LUV, NSC, RCL
49	Warehousing	2	UPS, FDX
51	Information	3	TV, VZ, DIS
52	Fin and insurance	17	USB, MTB, MCO, MMC, KEY MS, GS, BAC, C, AIG, AXP JPM, COF, WFC, HIG, PNC, CI
53	Real estate, rental/leasing	3	NLY, EQR, BXP
54	Prof, scientific and tech services	2	FLR, IBM
62	Health care	1	THC
72	Accommodation and food services	1	MCD

This table lists ticker symbols of our sample of 89 companies listed at the S&P 500 index during the period January 2, 2001, until December 31, 2019. All Tickers are grouped per sector according to the 2017 NAICS classification.

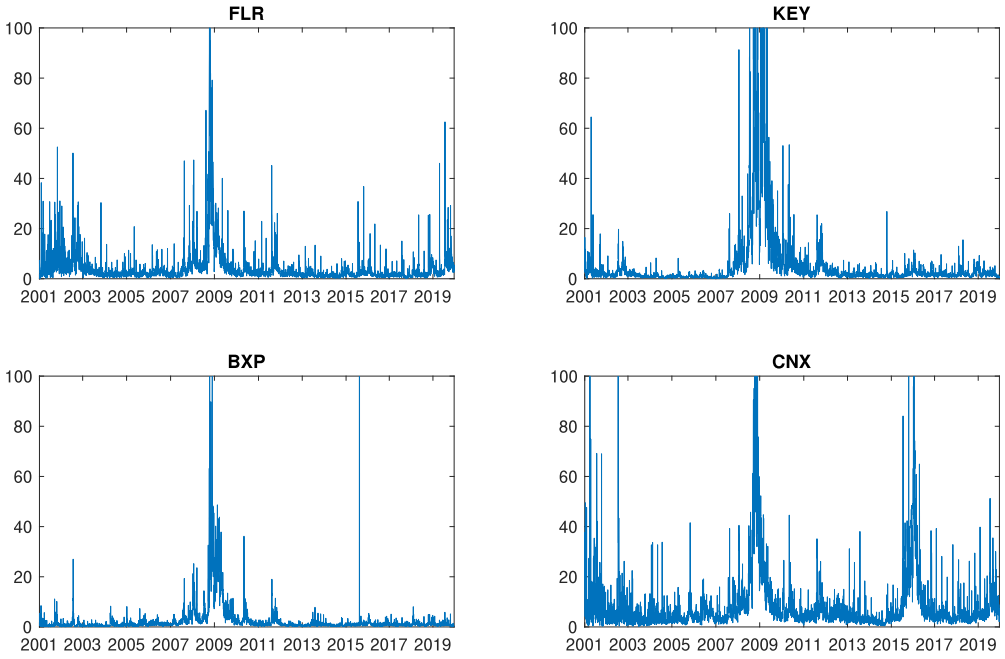

Fig. 4. Realized kernel time series. This figure shows the time series of the realized kernels (RK) of Fluor Corp, KeyCorp, Boston Properties INC, and Consol Energy Inc. For visual purposes, the vertical axes have been capped. The sample covers the period January 2, 2001, until December 31, 2019, and contains 4713 observations.

Table 2 shows summary statistics of the estimated parameters using all observations for all 89 individual stocks and the models for RK_t (Panel A) and $\log RK_t$ (Panel B), respectively. We report the means, standard deviations, and the 5th and 95th quantiles across all assets. Panel A shows the well-known high persistence in RK_t, which is measured by $\sum_{j=1}^{3} \beta_{\mu,j}$ in case of the GAS-HAR models, and by $\alpha_\mu + \sum_{j=1}^{3} \beta_{\mu,j}$ for the MEM-HAR model. The results also show strong persistence in the time-varying parameters $\nu_{1,t}$ and $\nu_{2,t}$, with the persistence being particularly strong for the skewness related parameter $\nu_{2,t}$. The parameter $\nu_{1,t}$, which affects more the vol-of-vol, is slightly less persistent. Panel B confirms the high persistence in $\log RK_t$. In addition, the vol-of-vol persistence (β_σ) is somewhat lower than the persistence of the variance itself ($\sum_{j=1}^{3} \beta_{\mu,j}$). This corresponds with panel

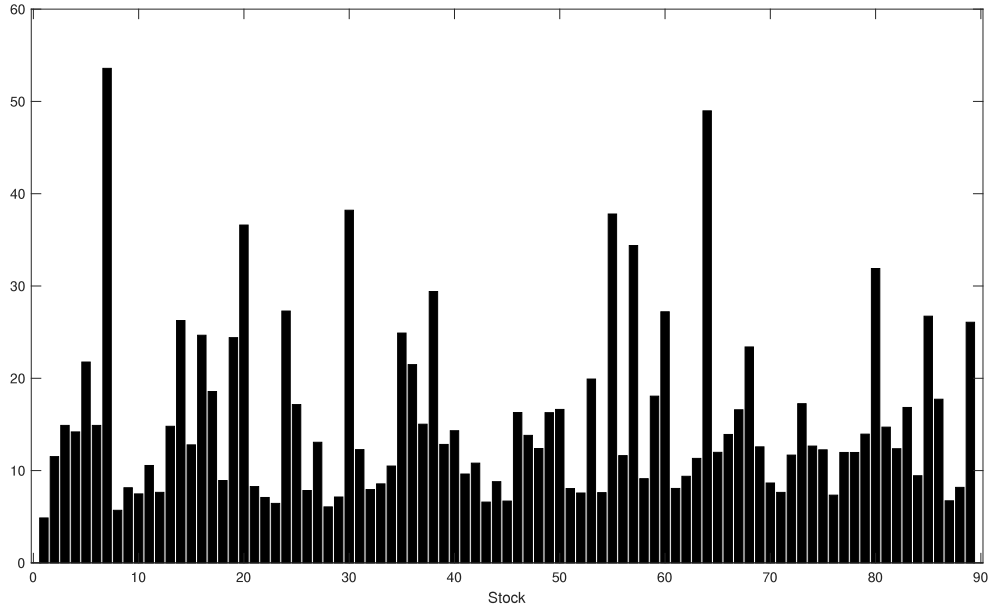

Fig. 5. Sample skewness of 89 stocks. This figure plots the sample skewness of daily realized kernels (RK) of 89 stocks listed at the S&P 500 Index. The sample covers the period January 2, 2001, until December 31, 2019, and contains 4713 observations.

A's slightly lower estimates of β_{ν_1} of the GAS-HAR-ν_1 model.

Fig. 6 shows differences in the Akaike Criterion, combining the number of estimated parameters with the maximized log-likelihood. A positive number means that the first-mentioned model has a better statistical fit corrected for the number of parameters. The four sub-panels have four clear messages. First, the large values in the upper left graph suggest that the conditional F distribution of the GAS-HAR model fits the Realized Kernel considerably better than the Gamma distribution of the MEM-HAR model. Second, allowing for time-varying DoF parameters in general increases the statistical fit versus the static GAS-HAR model, as indicated by the upper-right graph. Third, the bottom-left figure suggests that allowing for a time-varying vol-of-vol parameter within model specifications of the log volatility increases the statistical fit for each stock. Finally, the bottom-right figure indicates that for most stocks modeling the realized kernel itself with time-varying DoF parameters has better AIC values than modeling the logarithm of the realized kernel with a time-varying vol-of-vol.

Fig. 7 presents fitted values of $\nu_{1,t}$ and $\nu_{2,t}$ for the GAS-HAR-ν_{1-2} model for 11 stocks: PXD, KEY, EQR, JCP, AIG, NOC, SYY, IR, FDX, GIS and GD. The left panel shows the vol-of-vol path, and the right panel depicts the skewness paths, both as implied by the time-varying $\nu_{1,t}$ and $\nu_{2,t}$. The figure shows that indeed the vol-of-vol or dispersion varies through time, confirming the earlier found results of Corsi et al. (2008). For example, the vol-of-vol of AIG became very large during the crisis period 2008–2009. In addition, the lower-left panel shows quite some commonality in the VoV of certain stocks. The VoV and the skewness show time-varying behavior, as suggested by the sub-graphs of the right panel. The skewness is particularly high during turbulent periods, such as the 2001 US recession and/or the 2013 Sovereign Debt Crisis. It may even increase out of bounds if ν_2 approaches or goes below six during certain periods. It is also interesting to see increases in skewness during the Global Financial Crisis at the end of 2008 for some stocks, though not for others. Also, during the recovery in 2009, we saw different patterns across the 11 stocks. Allowing both parameters to vary thus appears empirically important. The following section investigates the possible implications of these patterns in an out-of-sample analysis.

Based on the full-sample analysis, we conclude that the vol-of-vol and tail-shape parameters of the distribution of the realized kernel series vary through time. Allowing for this time variation improves the fit of the conditional distribution. The following subsection investigates whether allowing for time-varying DoF parameters persists out-of-sample.

3.3. Out-of-sample results

We assess the short-term forecasting performance of the models in an economic setting by predicting 1-step density forecasts. In addition, we focus on the 1-step ahead Volatility-at-Risk (VolaR), a specific quantile of the density of the realized kernel. As indicated by Caporin et al. (2017), the VolaR is important for volatility traders and hedgers. We use the same set of models as for the in-sample analysis. We use a recursive estimation approach for each model starting from an initial sample period of 1500 observations. This covers the period 2001–2006, just before the start of the global financial crisis, and therefore constitutes a strong test on the forecasting performance of all models. We update the recursive estimates roughly every two months (50 observations).

We test the adequacy of our 1-step ahead density forecasts in two ways. The first test is based on the log

Table 2
Full-sample parameter estimates.

Model		ω_μ	α_μ	$\beta_{\mu,1}$	$\beta_{\mu,2}$	$\beta_{\mu,3}$	ν_1	ν_2	$\bar{f}_1$	α_{ν_1}	β_{ν_1}	$\bar{f}_2$	α_{ν_2}	β_{ν_2}
Panel A: modeling RK_t														
GAS ν_{1-2}	mean	0.041	0.954	0.844	0.097	0.037			2.848	0.240	0.859	2.506	0.039	0.961
	std	(0.024)	(0.081)	(0.022)	(0.022)	(0.011)			(0.391)	(0.182)	(0.171)	(0.413)	(0.036)	(0.083)
	5%	0.020	0.769	0.802	0.065	0.021			2.043	0.015	0.473	1.463	0.000	0.848
	95%	0.103	1.000	0.875	0.139	0.056			3.215	0.546	0.999	2.959	0.103	1.000
GAS ν_1	mean	0.041	0.933	0.843	0.100	0.035		16.49	2.748	0.196	0.871			
	std	(0.027)	(0.096)	(0.022)	(0.022)	(0.011)		(3.63)	(0.517)	(0.146)	(0.168)			
	5%	0.019	0.730	0.806	0.068	0.019		11.24	1.722	0.034	0.489			
	95%	0.105	1.000	0.875	0.144	0.054		21.58	3.199	0.442	0.999			
GAS ν_2	mean	0.038	0.960	0.851	0.093	0.038	21.45					2.498	0.051	0.927
	std	(0.021)	(0.073)	(0.024)	(0.022)	(0.011)	(3.51)					(0.477)	(0.032)	(0.189)
	5%	0.018	0.803	0.802	0.064	0.023	15.92					1.317	0.009	0.678
	95%	0.082	1.000	0.885	0.135	0.058	26.41					2.968	0.115	1.000
GAS	mean	0.038	0.946	0.846	0.100	0.037	20.69	15.76						
	std	(0.024)	(0.085)	(0.025)	(0.023)	(0.011)	(3.84)	(3.15)						
	5%	0.017	0.762	0.799	0.069	0.022	14.81	11.00						
	95%	0.096	1.000	0.879	0.146	0.056	26.01	20.84						
MEM H	mean	0.105	0.508	0.232	0.139	0.091	3.670							
	std	(0.074)	(0.078)	(0.078)	(0.052)	(0.042)	(0.737)							
	5%	0.041	0.392	0.116	0.041	0.031	2.540							
	95%	0.228	0.642	0.356	0.219	0.185	4.932							
Panel B: modeling $\log RK_t$														
model		ω	α_1	$\beta_{1,1}$	$\beta_{1,2}$	$\beta_{1,3}$	σ^2	ω_2	α_2	β_2				
GAS log-N VoV	mean	0.005	0.413	0.802	0.140	0.042		0.052	0.047	0.786				
	std	(0.009)	(0.038)	(0.031)	(0.030)	(0.012)		(0.052)	(0.030)	(0.212)				
	5%	−0.006	0.364	0.741	0.100	0.026		0.001	0.005	0.430				
	95%	0.023	0.482	0.846	0.192	0.061		0.164	0.098	0.997				
GAS log-N	mean	0.006	0.419	0.800	0.145	0.040	0.252							
	std	(0.008)	(0.045)	(0.034)	(0.032)	(0.012)	(0.047)							
	5%	−0.005	0.359	0.726	0.101	0.023	0.192							
	95%	0.023	0.505	0.843	0.208	0.059	0.330							

This table reports summary statistics of maximum likelihood parameter estimates of several models for the level and logarithm of daily realized kernels of 89 stocks from the S&P 500 index. Panel A lists results of modeling RK_t using the GAS-HAR models with ν_1 and ν_2 both time-varying (GAS ν_{1-2}), only ν_1 or ν_2 time-varying (GAS ν_1 and GAS ν_2), and both shape parameters static (GAS), and the MEM-HAR model (MEM). Panel B shows parameter estimates of the GAS-HAR-log-N with dynamic (GAS log-*N*-VoV) or static (GAS log-*N*) vol-of-vol parameters, applied to the logarithm of RK_t. We show the mean, standard deviation and the 5% and 95% quantile of the parameters across all 89 stocks. The sample covers the period January 2, 2001, until December 31, 2019 and contains 4713 observations.

score (see Amisano & Giacomini, 2007; Mitchell & Hall, 2005) as a scoring rule to differentiate between the density forecasts of the models. The log score of model M_j at time t of the density forecast is given by

$$S_{ls,t}(RK_t, M_j) = \log p_t(RK_t|\mathcal{F}_{t-1}, M_j) \tag{11}$$

for $t = 1501, 1502, \ldots, T-1$, where $p_t(\cdot)$ denotes the conditional density function of model M_j. Since we have quite some models under consideration, we account for possible interdependence between all density forecasts by using the Model Confidence Set approach of Hansen et al. (2011), applied on (minus) the log score of all models and using a significance level of 5%.[4]

Second, we also test the absolute performance of the density forecasts using the Berkowitz (2001) test, which is based on the Probability Transform Integral (PIT)

$$u_t(M_j) = F_t(RK_t|\mu_t, \mathcal{F}_{t-1}, M_j), \tag{12}$$

with $F_t(\cdot)$ the CDF of the distribution of model M_j. Instead of directly testing whether $u_t(M_j)$ is uniformly distributed, the test is based on $z_t(M_j) = \Phi^{-1}\big(u_t(M_j)\big)$, which should be N(0,1) and i.i.d. distributed under the null hypothesis of correct conditional coverage. Berkowitz (2001) considers a first-order autoregressive alternative with mean and variance possibly different from 0 and 1, respectively:

$$z_t(M_j) - c = \rho \cdot (z_{t-1}(M_j) - c) + \epsilon_t, \tag{13}$$

such that the null hypothesis boils down to $c = \rho = 0$ and $\mathrm{Var}(\epsilon_t) = 1$. The Likelihood Ratio test statistic now reads

$$LR = -2\big(L(0,1,0) - L(\hat{c}, \hat{\sigma}^2, \hat{\rho})\big), \tag{14}$$

with $L(\cdot)$ the log-likelihood. Under the $H(0)$, this test statistic is asymptotically χ^2 distributed with 3 degrees of freedom.

It is important to note that the above test focuses *jointly* on mean zero, variance one *and* serial independence of z_t. As we will see later, serial dependence may be

[4] Strictly speaking, a MCS is designed based on a moving (rather than expanding) window scheme. However, as pointed out in footnote 12 of Hansen et al. (2011), a recursive window scheme produces results that are very similar. As a robustness check, we also did all analyses using a moving window with a window length of 1500 observations. The main results do not change but become somewhat less strong than the recursive window scheme. We thank the referee for this point.

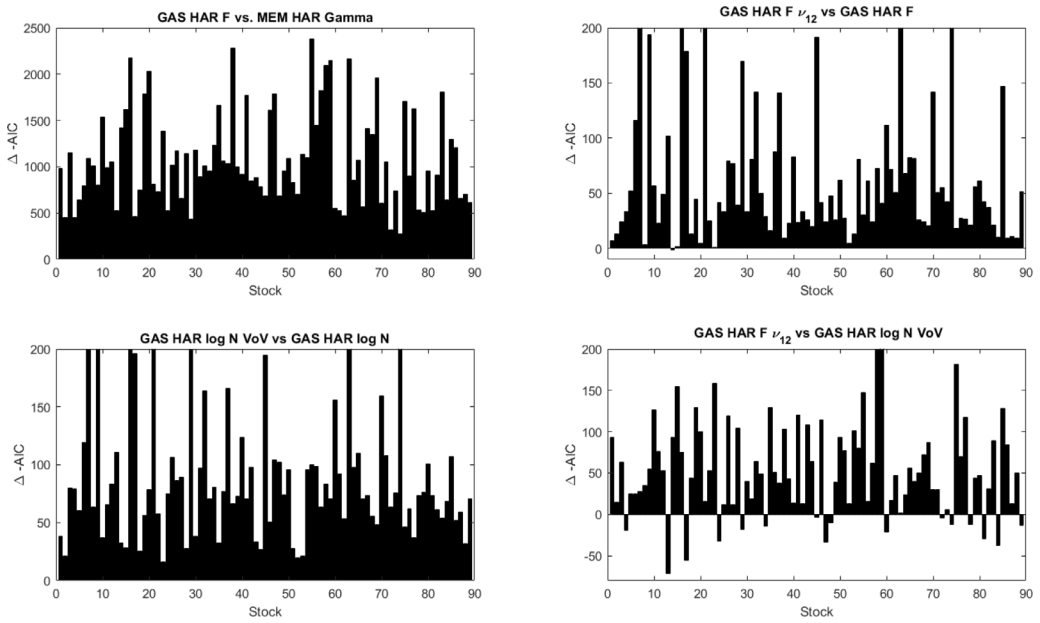

Fig. 6. Differences in AIC values. This figure shows the differences of the Akaike Information Criterion (AIC) between several score-driven and benchmark models of the dynamics of (log) RK_t. A positive number means that the first mentioned model has a lower AIC. The sample covers the period January 2, 2001, until December 31, 2019, and contains 4713 observations.

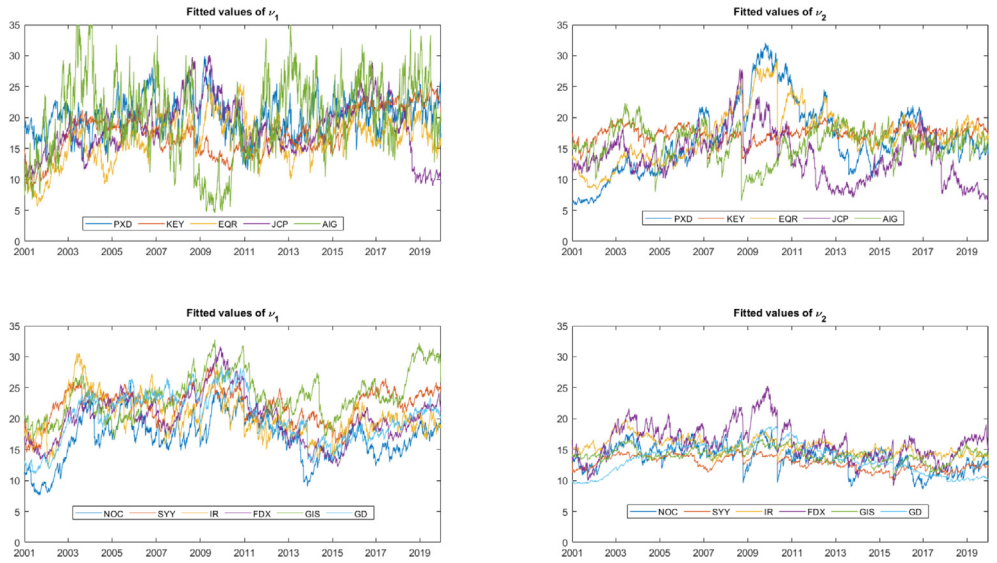

Fig. 7. Fitted values of the D.o.F. parameters. This figure shows the fitted time-varying values of $\nu_{1,t}$ and $\nu_{2,t}$ associated with the eleven stocks according to the GAS-HAR-ν_{1-2} model. The upper panels show PXD, KEY, EQR, JCP, and AIG results, while the lower panel depicts fitted DoFs of NOC, SYY, IR, FDX, GIS, and GD. The sample covers the period January 2, 2001 until December 31 2019 and contains 4,312 observations.

challenging for all models considered. We therefore also consider the 'unconditional' Berkowitz test focusing only on c and σ^2 by calculating

$$LR_{SN} = -2\big(L(0,1,\hat{\rho}) - L(\hat{c},\hat{\sigma}^2,\hat{\rho})\big), \tag{15}$$

which is asymptotically $\chi^2(2)$ distributed under the null of $H_0 : c = 0, \sigma^2 = 1$.

As a second empirical out-of-sample application, we consider the 1-step ahead VolaR. The VolaR computes the risk of extremely high volatility. The q% VolaR is defined as the q-quantile of the distribution of the realized kernel,

$$P[RK_{t+1} > VolaR^q_{t+1}|\mathcal{F}_t] = q, \tag{16}$$

where q is set to 5 percent. Note that the probability depends both on μ_t and on the shape parameters of the distribution used. We backtest our predicted *VolaR* using the unconditional and conditional coverage (UC and CC) tests proposed by Christoffersen (1998). The former tests whether number of violations, i.e. the number of times

Table 3
1-step ahead density and Volatility-at-Risk predictions.

	RK_t					$\log RK_t$	
	GAS ν_{1-2}	GAS ν_1	GAS ν_2	GAS	MEM	GAS log-N-VoV	GAS log-N
Panel A: Unconditional backtest results							
Panel A.1: log score: # times (out of 89) the model is in the 95% model confidence set (MCS)							
♯ in MCS	**87**	82	83	74	11	66	42
Panel A.2: tick loss: # times (out of 89) the model is in the 95% model confidence set (MCS)							
♯ in MCS	58	**69**	32	42	29	66	47
Panel A.3: # stocks with Unconditional Coverage test rejections							
♯ p-val < 0.10	23	38	**16**	42	41	34	42
♯ p-val < 0.05	**14**	26	**14**	35	38	26	32
♯ p-val < 0.01	7	13	**5**	20	25	12	22
Panel A.4: # stocks with Unconditional Berkowitz tests (excluding serial dependence) rejections							
♯ p-val < 0.10	**42**	43	60	64	77	49	61
♯ p-val < 0.05	**30**	38	52	61	76	42	58
♯ p-val < 0.01	**20**	29	37	52	68	34	51
Panel B: Conditional backtest results							
Panel B.1: # stocks with Conditional Coverage test rejections							
♯ p-val < 0.10	88	89	85	89	**48**	69	83
♯ p-val < 0.05	88	89	84	88	**40**	61	80
♯ p-val < 0.01	87	89	81	85	**27**	39	63
Panel B.2: # stocks with Berkowitz tests rejections							
♯ p-val < 0.10	89	88	89	89	**81**	83	87
♯ p-val < 0.05	89	88	89	88	**79**	80	83
♯ p-val < 0.01	87	88	86	87	**67**	77	79

This table reports one 1-step ahead density and 95% Volatility-at-Risk predictions of the realized kernel. We consider the MEM-HAR model assuming a Gamma distribution, the GAS-HAR-F model assuming an F distribution with fixed degrees of freedom parameters, two GAS-HAR models with time-varying shape and/or dispersion parameter(s), and a GAS model for the logarithm of the realized kernel with and without time-varying volatility-of-volatility (denoted as GAS log-N (VOV)). The models are applied to daily realized kernels of 89 stocks from the S&P 500 index. We use an expanding window where the first window contains 1500 observations. Panel A summarizes the results of unconditional backtests, while Panel B focuses on conditional backtests. More specifically, Panels A.1 and A.2 denote the number of times a model belongs to the model confidence set (MCS), applied on the log score and tick loss, respectively, using a 5% significance level. Panel A.3 lists the number of times the p-value associated with the null-hypothesis of the Christoffersen (1998) tests on unconditional coverage of the 95% VolaR respectively is below 10, 5, or 1%. Panel A.4 summarizes the number of times the p-value of the unconditional Berkowitz test on correct absolute density forecasts excluding serial dependence for the inverse standard normal transformation of the PITs falls below the 10, 5, or 1% level. Panel B shows the results of two conditional backtests. The number of times the p-value associated with the null-hypothesis of the Christoffersen (1998) tests on conditional coverage of the VolaR predictions respectively is below 10, 5 or 1% is shown by panel B.1. Finally, Panel B.2 shows results on the Berkowitz test for the transformed PITs, but now including possible serial dependence. A bold number represents the best model, i.e., the model that most often belongs to the set (panel A.1) or the model with the lowest number of rejections in each row (remaining panels). The out-of-sample period is from January 17, 2007 until December 31 2019 and contains 3213 observations.

$RK_{t+1} > VolaR^q_{t+1}$, equals the unconditional coverage probability q, while the latter in addition tests for clustering in the VolaR violations. Finally, we compare the relative performance of the VolaR estimates of all models using the following asymmetric linear tick loss function, which is, for example, also used in the conditional predictive ability (CPA) test of Giacomini and White (2006):

$$L^q(e_{t+1}) = (q - I[e_{t+1} < 0])e_{t+1}, \quad e_{t+1} = VolaR^q_{t+1} - RK_{t+1} \tag{17}$$

This loss function is asymmetric in the sense that if a VolaR violation occurs, the loss equals $(1-q)e_{t+1}$, while in case of no violation, the loss equals qe_{t+1}, which is in turn considerably lower. Similar to the unconditional density forecasts, we use the MCS to compare the losses of all models.

Table 3 shows all out-of-sample results. Panel A lists results relating to the *unconditional* accuracy of the models, whereas Panel B focuses on the *conditional* performance.

To take the full distributional properties of the forecasts into account, Panel A.1 in Table 3 shows results for the relative performance of the 1-step ahead density forecasts using the Model Confidence Set approach of Hansen et al. (2011) with a significance level of 5%. The GAS-HAR-ν_{1-2} model performs best: it almost always belongs to the model confidence set with 87 out of 89 cases. On the other hand, the MEM-HAR model only enters the set 11 times. Overall, the F distribution (and hence all GAS-HAR type models) appears to produce considerably better density forecasts than the Gamma distribution. Comparing the static F distribution (GAS) with the dynamic F distribution (GAS-HAR-$\nu_.$), the improvement effect is still clearly noticeable: the static model enters the MCS 74 times, compared to the 87 of the model with both parameters time-varying. We also see that the models using the logarithm of the realized kernel, a normal distribution, and time-varying volatility enter the MCS 66 times, which is fair, but still far below the 87 of the GAS-HAR-ν_{1-2} model. The model for $\log RK_t$ with static volatility fares even worse with an entry of 42.

Panel A.2 again shows MCS results using a 5% significance level, but now applied to the tick loss function to compare the relative performance of the 95% VolaR forecasts. The GAS-HAR-ν_1 model performs best (69 out of 89 cases it belongs to the model confidence set), closely followed by the GAS log-N-VoV and GAS-HAR-ν_{1-2} models. Put differently, time-varying shape parameters (either in the kernel or logarithm of the kernel) seem to be important. Note that the difference between modeling the realized kernel or its logarithm is less pronounced. Moreover, the MEM-HAR model again does not perform well as it only belongs to the model confidence set 29 times.

Panel A.3 focuses on the 95% VolaR, a specific quantile of the distribution of the realized kernel. Two important conclusions can be drawn. First, allowing for a time-varying $\nu_{2,t}$ produces by far the best unconditional VolaR predictions, followed by a model with time-varying $(\nu_{1,t}, \nu_{2,t})$. The differences between the GAS-HAR-ν_2 and ν_{1-2} models versus the GAS-HAR model are striking: the number of rejected null-hypotheses consistently exceeds a factor 2 (for instance, 42 vs. 16, 35 vs. 14 and 20 vs. 5). This holds even more strongly if we compare the same model against the MEM-HAR models. Moreover, only allowing the 'vol-of-vol parameter' $\nu_{1,t}$ to be time-varying improves the VolaR predictions to some extent, but we still reject the null hypothesis of correct unconditional coverage in 26 (12) cases at a 5% (1%) significance level. Hence allowing for a time-varying $\nu_{2,t}$ parameter appears crucial. Second, the unconditional coverage results show that it is better to model the realized kernel distribution directly (using a fat-tailed distribution with time-varying shape parameters) than to model the distribution of the log-kernel $\log RK_t$ using a normal distribution with time-varying volatility. Allowing for a time-varying VoV improves the VolaR forecasts compared to a fixed VoV parameter for the log-kernel models. Still, we reject the null-hypothesis of unconditional coverage 34, 26, and 12 times, respectively, which is 80% to more than 100% times more than for the GAS-HAR-ν_2 model.

Finally, Panel A.4 looks at the absolute density forecasts using the 'unconditional' Berkowitz test. Although the GAS-HAR-ν_{1-2} again outperforms its competitors, the result is less strong here: in 30 out of 89 cases, the null hypothesis has been rejected using a 5% significance level. This leads to the important result that it is, in general, somewhat challenging to predict accurate unconditional density forecasts over the entire tail area. We shed more light on this result when discussing the conditional Berkowitz test below.

Panel B of Table 3 provides a summary of the conditional test results. Both panels convey the same message: all models have difficulties producing accurate conditional VolaR predictions. The only extra signal from Panel B appears that the MEM-HAR model fare (somewhat) better than the GAS-HAR models in Panel B.1. This turns out to be mainly due to the MEM-HAR model exhibiting less serial dependence. The challenge GAS models face with serial dependence of the violations is not entirely surprising and partly inherent to the use of the score-driven modeling methodology. Remember the score in, for instance, (6) downweights large observations in a trade-off of whether such observations are caused by the fat-tailedness of the data or by recently increased volatility. Only if more large observations occur after an initial one, the score-driven volatility increases as it becomes clear that such a sequence of large observations cannot be attributed to fat-tailedness alone. This induces some of the correlations that cause the outcome of the CC tests for GAS models. Investigating this further, we also find that these correlations for the GAS models are very short-lived. Second order partial autocorrelations are generally very close to zero for all models, indicating the GAS score-driven models only require a few signals to disentangle fat-tailedness from increases in volatility.

We also scrutinize the conditional results for the MEM-HAR model a bit further in Fig. 8 to provide further insights into the difference between the MEM-HAR and GAS-HAR-ν_{1-2} conditional results. We show scatter plots of z_t against z_{t-1} for three different representative stocks. Under the null hypothesis, z_t and z_{t-1} should be uncorrelated and have a bivariate standard normal distribution. In all three cases, the MEM-HAR specification appears to be plagued by quite a few severe horizontal and vertical influential observations, thus severely violating the bivariate normality. The pattern is in line with that of additive outliers (AOs) in a time series setting (Martin & Yohai, 1986, see for instance), which are known to cause a downward bias in estimates of serial dependence parameters such as $\hat{\rho}$ in a subsequent increase in the number of 'non-rejections' for the MEM-HAR model of conditional tests based on such serial correlation estimates. This further explains the better behavior of the MEM-HAR model in Panel B (conditional) combined with the worse behavior in Panel A (unconditional). By contrast, the properties of z_t in the right-hand panels for the GAS-HAR models appear closer to normality than the left-hand panels: univariate Bera-Jarque tests for non-normality are a 40-fold higher for the MEM-HAR model than for the GAS-HAR models.

We can conclude that unconditionally, the new GAS-HAR models with an F distribution and time-varying parameters perform best, both in terms of model confidence sets, (modified) Berkowitz tests, and unconditional coverage tests. Conditionally, all models still face substantial challenges. Part of these are inherent to the methodology: the GAS models need more than a single signal to disentangle fat-tailedness from volatility increases. However, the induced serial correlations in the diagnostics are short-lived, though this does not show up in the typical numerical test results. Conversely, challenges and serial correlations faced by non-robust models may be masked by typically reported numerical test results, as we saw in Fig. 8 for the MEM-HAR model compared to the new GAS-HAR model. It, therefore, seems a good idea to constantly scrutinize numerical test results further, possibly with additional graphical analyses.

4. Conclusions

We introduced a new dynamic score-driven model for the vol-of-vol and skew-of-vol of realized kernels. The proposed model explicitly acknowledges that realized kernels are fat-tailed and have robust propagation

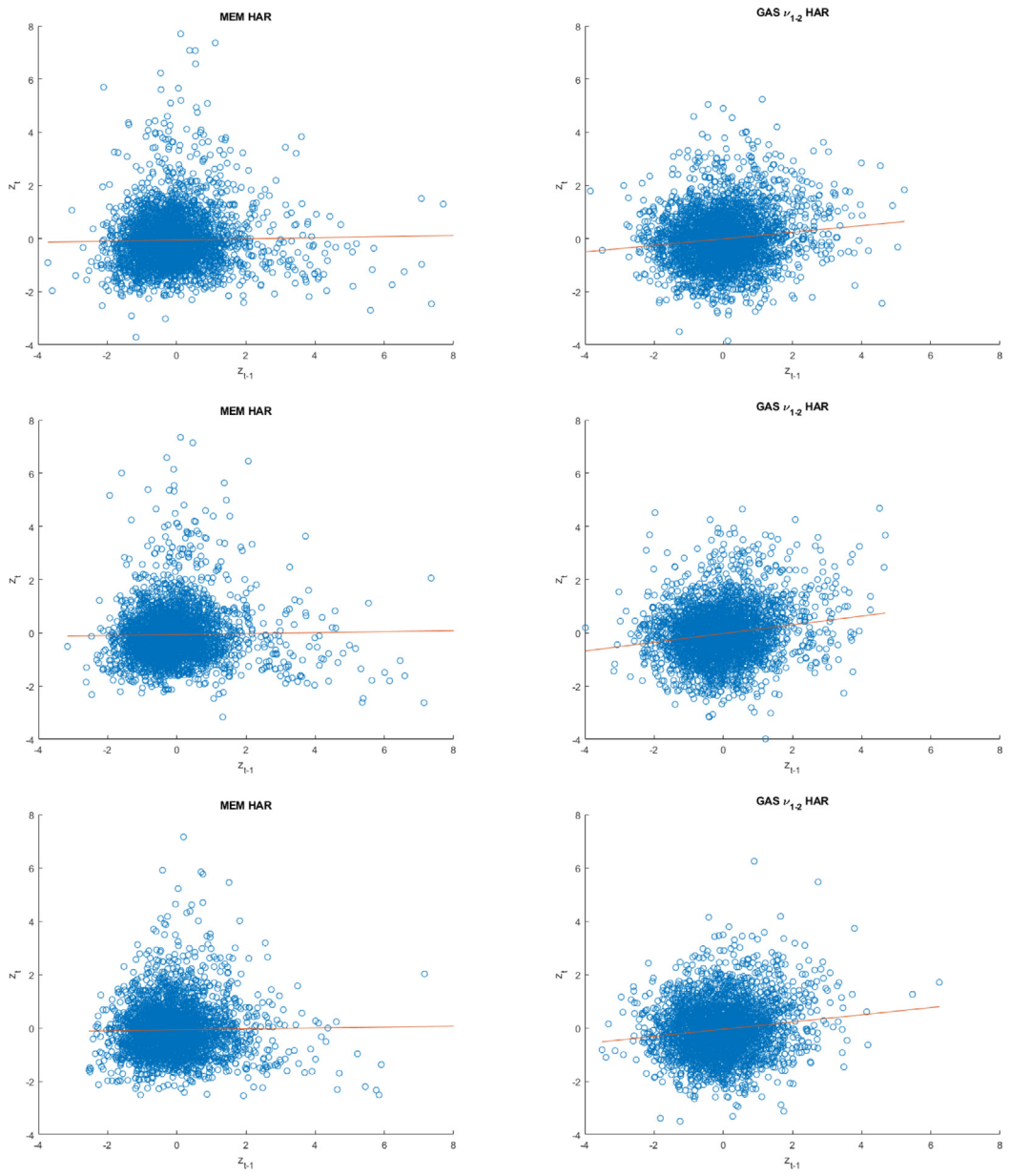

Fig. 8. Scatter plots of transformed PITs. This figure shows a scatter plot of the transformed PITs z_t against z_{t-1} for the stocks LLY, MCO and HON. The left three graphs are associated with the MEM-HAR model, the right three graphs correspond to the GAS-HAR-ν_{1-2} model. The out-of-sample period covers January 17, 2007, until December 31, 2019 and contains 3213 observations.

dynamics for the time-varying parameters. The proposed setup is particularly suitable for cases where no explicit robustification methods are applied while estimating realized measures. Using realized kernels of 89 US stocks over 2001–2019, the new model improves both the in-sample and out-of-sample fit of the realized kernel dynamics vis-á-vis the MEM model (with HAR dynamics) of Engle and Gallo (2006) and the GAS F of Opschoor et al. (2018) model with static parameters.

Summarizing the out-of-sample Volatility-at-Risk (VolaR) and density prediction results, we conclude that unconditional density and VolaR forecasts clearly improve when accounting for time-variation in both $\nu_{1,t}$ and $\nu_{2,t}$ and thus in the vol-of-vol and skew-of-vol. Gathering all results simultaneously in a model-confidence-set (MCS) approach, we see that the new model is almost always (87 out of 89) part of the MCS. The same conclusion holds with respect to forecasting the VolaR: the new model has better unconditional coverage properties. In addition, we see that modeling the realized kernel directly with the time-varying F distribution as suggested in our paper produces better VolaR and density forecasts than modeling the logarithm of RK_t, as is regularly proposed in the literature.

All models face challenges for conditional forecasts. Interestingly, though, better numerical test results for

serial dependence for the MEM-HAR model may mask the occurrence of sequences of influential observations in the PITs of these models, possibly causing under-rejections of the test similar to additive outlier problems in a time-series context. In particular, the failure of the normality seems much more pronounced for the MEM-HAR model than for the GAS-HAR models. Also, the serial correlations of the GAS approach are very short-lived and partly attributable to the methodology requiring more than one signal to disentangle a fat-tailed observation from a large observation due to a volatility increase. This calls for the further development of tests that are robust to the problem of influential observations and, at the same time, reflect any failure of the bivariate normality under the null of the correct specification. Nevertheless, given our unconditional backtest results, we conclude that the new model provided a valuable tool when modeling and forecasting realized kernels; including time-varying degrees-of-freedom parameters is empirically important.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

Amisano, G., & Giacomini, R. (2007). Comparing density forecasts via weighted likelihood ratio tests. *Journal of Business & Economic Statistics*, *25*, 177–190.

Andersen, T. G. (1996). Return volatility and trading volume: An information flow interpretation of stochastic volatility. *The Journal of Finance*, *51*, 169–204.

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

Audrino, F., & Knaus, S. D. (2016). Lassoing the HAR model: A model selection perspective on realized volatility dynamics. *Econometric Reviews*, *35*, 1485–1521.

Baltussen, G., Van Bekkum, S., & Van Der Grient, B. (2018). Unknown unknowns: uncertainty about risk and stock returns. *Journal of Financial and Quantitative Analysis*, *53*, 1615–1651.

Barndorff-Nielsen, O. E., Hansen, P. R., Lunde, A., & Shephard, N. (2008). Designing realized kernels to measure the ex post variation of equity prices in the presence of noise. *Econometrica*, *76*, 1481–1536.

Barndorff-Nielsen, O. E., Hansen, P. R., Lunde, A., & Shephard, N. (2009). Realized kernels in practice: trades and quotes. *The Econometrics Journal*, *12*, 1–32.

Berkowitz, J. (2001). Testing density forecasts, with applications to risk management. *Journal of Business & Economic Statistics*, *19*, 465–474.

Blasques, F., Koopman, S. J., & Lucas, A. (2015). Information theoretic optimality of observation driven time series models for continuous responses. *Biometrika*, *102*, 325–343.

Blasques, F., van Brummelen, J., Koopman, S. J., & Lucas, A. (2022). Maximum likelihood estimation for generalized autoregressive score models. *Journal of Econometrics*, *227*, 325–346.

Brownlees, C. T., & Gallo, G. M. (2006). Financial econometric analysis at ultra-high frequency: Data handling concerns. *Computational Statistics & Data Analysis*, *51*, 2232–2245.

Caporin, M., Rossi, E., & de Magistris, P. S. (2017). Chasing volatility: A persistent multiplicative error model with jumps. *Journal of Econometrics*, *198*, 122–145.

Chiriac, R., & Voev, V. (2011). Modelling and forecasting multivariate realized volatility. *Journal of Applied Econometrics*, *26*, 922–947.

Christoffersen, P. (1998). Evaluating interval forecasts. *International Economic Review*, *39*, 841–862.

Corradi, V., Distaso, W., & Mele, A. (2013). Macroeconomic determinants of stock volatility and volatility premiums. *Journal of Monetary Economics*, *60*, 203–220.

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

Corsi, F., Mittnik, S., Pigorsch, C., & Pigorsch, U. (2008). The volatility of realized volatility. *Econometric Reviews*, *27*, 46–78.

Cox, D. R. (1981). Statistical analysis of time series: some recent developments. *Scandinavian Journal of Statistics*, *8*, 93–115.

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

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

Gerlach, R., Lu, Z., & Huang, H. (2013). Exponentially smoothing the skewed Laplace distribution for value-at-risk forecasting. *Journal of Forecasting*, *32*, 534–550.

Giacomini, R., & White, H. (2006). Tests of conditional predictive ability. *Econometrica*, *74*, 1545–1578.

Golosnoy, V., Gribisch, B., & Liesenfeld, R. (2012). The conditional autoregressive Wishart model for multivariate stock market volatility. *Journal of Econometrics*, *167*, 211–223.

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

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

Huang, D., Schlag, C., Shaliastovich, I., & Thimme, J. (2019). Volatility-of-volatility risk. *Journal of Financial and Quantitative Analysis*, *54*, 2423–2452.

Jin, X., & Maheu, J. M. (2016). Bayesian semiparametric modeling of realized covariance matrices. *Journal of Econometrics*, *192*, 19–39.

Lucas, A., & Zhang, X. (2016). Score-driven exponentially weighted moving averages and value-at-risk forecasting. *International Journal of Forecasting*, *32*, 293–302.

Martin, R. Douglas, & Yohai, Victor J. (1986). Influence functionals for time series. *The Annals of Statistics*, 781–818.

Mitchell, J., & Hall, S. G. (2005). Evaluating, comparing and combining density forecasts using the KLIC with an application to the bank of England and NIESR fan charts of inflation. *Oxford Bulletin of Economics and Statistics*, *67*, 995–1033.

Noureldin, D., Shephard, N., & Sheppard, K. (2012). Multivariate high-frequency-based volatility (HEAVY) models. *Journal of Applied Econometrics*, *27*, 907–933.

Opschoor, A., Janus, P., Lucas, A., & van Dijk, D. (2018). New HEAVY models for fat-tailed realized covariances and returns. *Journal of Business & Economic Statistics*, *36*, 642–657.

Shephard, N., & Sheppard, K. (2010). Realising the future: forecasting with high-frequency-based volatility (HEAVY) models. *Journal of Applied Econometrics*, *25*, 197–231.

Sinclair, E. (2013). *Volatility trading*. John Wiley & Sons.